JOURNAL OF
ATMOSPHERIC AND TERRESTRIAL PHYSICS

EDITORIAL ADVISORY BOARD

H. O. G. ALFVÉN-STOCKHOLM · J. BARTELS-GÖTTINGEN · P. M. S. BLACKETT-MANCHESTER · E. G. BOWEN-SYDNEY · SIR DAVID BRUNT-LONDON · R. BUREAU-BAGNEUX · S. CHAPMAN-OXFORD · J. EGEDAL-COPENHAGEN · H. FICKER-VIENNA V. F. HESS-NEW YORK · E. O. HULBURT-WASHINGTON, D.C. · H. ISRAEL-TÜBINGEN SIR NELSON JOHNSON-LONDON · D. H. MENZEL-CAMBRIDGE, MASS. S. K. MITRA-CALCUTTA · M. NICOLET - UCCLE · E. H. NORINDER - UPSALA B. van der POL-GENEVA · J. A. RATCLIFFE - CAMBRIDGE E. REGENER-WEISSENAU (BODENSEE) O. E. H. RYDBECK-GOTHENBURG SIR GEORGE SIMPSON-LONDON · H. SOLBERG-OSLO · C. STØRMER-OSLO H. U. SVERDRUP-OSLO · J. VELDKAMP-DE BILT M. WALDMEIER-ZÜRICH · R. v. d. R. WOOLLEY-CANBERRA J. ZENNECK-ALTHEGNENBERG

EDITOR IN CHIEF

SIR EDWARD APPLETON, F.R.S.
EDINBURGH

VOLUME 4
1953

PERGAMON PRESS LTD. LONDON

L.
53

J. Atmosph. Terr. Phys.

VOL.
4
1953

Contents

Nos. 1/2

PAGE

I. C. Browne and T. R. Kaiser: The radio echo from the head of meteor trails 1

W. Petrie: Rotational temperatures of auroral nitrogen bands 5

Rune Lindquist: An interpretation of vertical incidence equivalent height versus time recordings on 150 Kc/s 10

S. Deb: Winds in the ionospheric regions 28

R. P. Waldo Lewis and D. H. McIntosh: Geomagnetic and ionospheric relationships . . 44

C. W. Allen: World wide diurnal vibration on the F_2 region 53

G. Elwert: Über die Bedeutung einer weichen Röntgenstrahlung der Sonnenkorona für dei Bildung der normalen ionosphärischen E-Schicht 68

Research Notes:

R. P. Waldo Lewis and D. H. McIntosh: A Universal time component in geomagnetic disturbance . . 78

N. C. Gerson: A note on auroral interaction 81

B. W. Osborne: The Electron content of the F_2 layer above Singapore 83

J. A. Chalmers: Induced changes and the fine-weather field 84

Indices of geomagnetic activity of the observatories at Abinger, Eskdalemuir and Lerwick, April to August 1953 86

Book reviews 90

No. 4

H. Norinder and R. Siksna: The mobility of atmospheric small ions during summer nights at Uppsala 93

R. Siksna: Mobility of small atmospheric ions in the air from the ground at Uppsala . . 106

D. R. Bates and A. Dalgarno: The altitudes of the luminous layers in the earth's atmosphere . 112

J. Alan Chalmers: The agrimeter for continuous recording of the atmospheric electric field . 124

F. Horner: The reception of atmospherics at high frequencies 129

Nos. 4/5

A. P. Mitra and R. E. Jones: Determination of the location of the ionospheric current system responsible for geomagnetic effects of solar flares 141

J. W. Dungey: Bailey's theory of sunspot noises 148

R. E. Loughhead: The equations for a problem arising in Dungey's investigation of Bailey's theory of sunspot noise 163

A. A. Weiss: Solar tides in the F_2 region from the study of night-time critical frequencies . 175

J. A. Fejer: Semidiurnal currents and electron drifts in the ionosphere 184

A. P. Mitra and C. A. Shain: The measurement of ionospheric absorbtion using observations of 18·3 Mc/s cosmic radio noise 204

Laurence A. Manning: The strength of meteoric echoes from dense columns 219

M. A. Ellison: The $H\alpha$ radiation from solar flares in relation to sudden enhancements of atmospherics on frequencies near 27 Kc/s 226

D. H. Shinn: Analysis of ionospheric records (ordinary ray) Part I 240

D. S. Robertson; D. T. Eddy and W. G. Elford: Measurements of winds in the upper atmosphere by means of drifting meteor trails I 255

W. G. Elford and D. S. Robertson: Measurements of winds in the upper atmosphere by means of drifting meteor trails II 271

No. 6

PAGE

M. J. SEATON: Excitation processes in the aurora and airglow 1 285

M. J. SEATON: Excitation processes in the aurora and airglow 2 295

L. G. SMITH and G. F. SCHILLING: The variation of electrical conductivity of air within sealed rooms 314

L. HARANG and B. LANDMARK: Radio echoes during aurorae and geomagnetic storms using 35 and 74 Mc/s waves simultaneously 322

Research Notes:

G. W. GRIFFING and A. L. STEWART: The effect of the horizontal extent of aurorae on the apparent luminosity-altitude curves 339

N. C. GERSON: A note on solar-terrestrial relationships 341

VICTOR F. HESS: Remarks on the article of J. Clay 343

Indices of geomagnetic activity of the Observatories at Abinger, Eskdalemuir and Lerwick, September to November 1953 346

Author Index to Volume 4

(B) denotes a book review

ALLEN, C. W. p. 53
BATES, D. R. p. 112
BROWNE, I. C. p. 1
CHALMERS, J. A. pp. 84, 124
DEB, S. p. 28
DELGARNO, A. p. 112
DUNGEY, J. W. p. 148
EDDY, D. T. p. 255
ELFORD, W. C. pp. 255, 271
ELLISON, M. A. p. 226
ELWERT, G. p. 68
FEJER, J. A. p. 184
FRISCH, O. R. pp. 90 (B), 91 (B)
GERSON, N. C. pp. 81, 341
GRIFFING, G. W. p. 339
HARTSHORN, L. p. 91 (B)
HESS, V. F. p. 343
HORNER, F. p. 129
JONES, R. E. p. 141
KAISER, T. R. p. 1
LANDMARK, B. p. 322
LINDQUIST, R. p. 10
LOUGHHEAD, R. E. p. 163
MANNING, L. A. p. 219
MCINTOSH, D. H. pp. 44, 78
MITRA, A. P. pp. 141, 204
NORINDER, H. p. 93
OSBORNE, B. W. p. 83
PETRIE, W. p. 5
ROBERTSON, D. S. pp. 255, 271
SCHILLING, G. F. p. 314
SEATON, M. J. pp. 285, 295
SHAIN, C. A. p. 204
SHINN, D. H. p. 240
SIKSNA, R. pp. 93, 106
SMITH, L. G. p. 314
SMITH-ROSE, R. L. p. 92 (B)
STEWART, A. L. p. 339
WALDO LEWIS, R. P. pp. 44, 78
WEISS, A. A. p. 175

VOL.
4
1953

Journal of Atmospheric and Terrestrial Physics, 1953. Vol. 4, pp. 1 to 4. Pergamon Press Ltd., London

The radio echo from the head of meteor trails

I. C. Browne and T. R. Kaiser

Jodrell Bank Experimental Station, University of Manchester

(*Received* 6 *May* 1953)

ABSTRACT

A theoretical account in terms of familiar diffraction theory is offered in explanation of a type of radio echo, apparently originating at the moving head of a meteor trail, which is sometimes observed in addition to the well-known specularly reflected echo. No assumptions are invoked beyond those commonly accepted as necessary to explain the speculary reflected echo. The theory predicts intensities of head echoes in agreement with observation, and also predicts that the intensity of the head echo relative to that of the specularly reflected echo should be proportional to the radio wavelength and inversely proportional to the difference in range between the two echoes. It is suggested that other anomalous meteor echoes may arise from discontinuities associated with visual flares.

It is well known that besides the specularly reflected, or "body," echo from a meteor trail, a short duration echo is sometimes seen, apparently originating at the head of a dense trail and moving with the velocity of the meteor [1], [2], [3], [4].

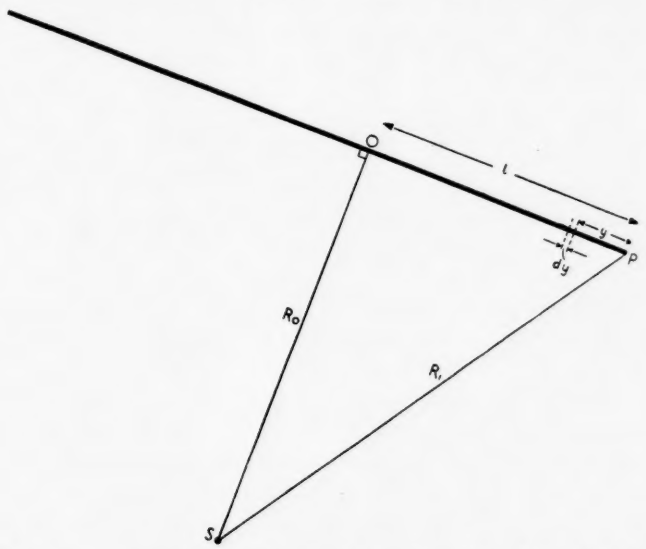


Fig. 1. The geometrical conditions for the formation of an echo from the head of a meteor trail

Its intensity increases with the radio wavelength. Millman and McKinley [1] have attributed this "head" echo to ionization produced by ultra-violet radiation from the incandescent meteor. An echo arising in this manner, however, should persist for longer than is usually observed, while it is difficult to see how sufficiently intense ultra-violet radiation could be produced by the meteor. The purpose of this note is to propose a simple theory to account for the head echo in terms of well-established properties of the ordinary meteor trail.

Suppose the meteor trail, at any instant, to consist of a narrow, semi-infinite column of ionization and to be illuminated by a radio pulse of duration T. Let the perpendicular from the observing station S (Fig. 1) meet the trail at the point O at range R_0, and let P be the instantaneous position of the meteor at range R_1. Then an elementary condition for the observation of a separate head echo is that $\frac{1}{2}cT < (R_1 - R_0)$, where c is the velocity of light.

At any instant, part of the trail will be illuminated, and an echo will be obtained only if: (a) the phase of the instantaneous oscillations of the electrons induced by the incident wave varies only slowly with distance along the trail (condition of stationary phase), or: (b) the reflection coefficient changes abruptly with distance along the trail.

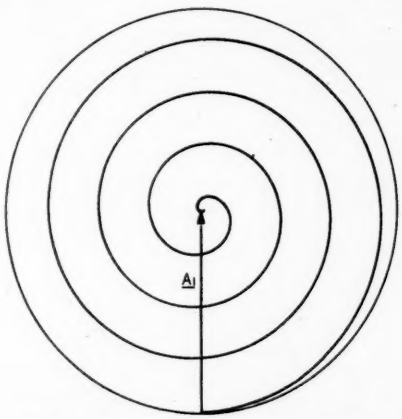


Fig. 2. Vector diagram representation of a head echo. The phase angle ϕ increases with clockwise rotation

The case (a) represents the condition under which the body echo is obtained from a region of the trail of length $\sim(R_0\lambda)^{\frac{1}{2}}$ (λ = wavelength) surrounding the point O.

Case (b) represents the condition for the return of an echo from the head of the trail, or from discontinuities which may arise at points on the trail.

The intensity of the head echo may be estimated as follows. When the meteor is at a distance l from O (Fig. 1) the wave scattered towards the receiver from an element of the trail distance y from P differs in phase from that scattered from P itself by ϕ, where

$$\phi = 4\pi ly/\lambda R_1$$

Thus ϕ varies linearly with y. If the amplitude of the wave scattered from the element of length dy be $F(y)dy$, the total amplitude of the head echo A_1 is given by,

$$A_1 = \left| \int_0^\infty F(y) \exp(j\phi) dy \right|$$

The form of $F(y)$ depends upon the shape of the transmitter pulse. Since the time of rise and decay of this pulse must occupy several radio-frequency cycles, $F(y)$

VOL. 4 1953

will fall to zero over several full period variations of ϕ. The integral in the expression for A_1 may thus be represented by the vector diagram of Fig. 2, A_1 being given by the closing vector of the spiral.

Hence,

$$A_1 = \tfrac{1}{2}F \left| \int_0^{\pi} (R_1\lambda/4\pi l) \exp(j\phi) d\phi \right| = R_1\lambda F/4\pi l$$

where $F = F(0)$ is the value corresponding to the peak of the transmitter pulse.

The amplitude of the body echo from a meteor trail of the same line density may be shown to be

$$A = F(R\lambda/2)^{\frac{1}{2}} \qquad \text{(Lovell and Clegg [5])}$$

Thus the ratio I_1/I of the intensity of the head echo to that of the body echo which would be obtained from a trail of the same line density at range R_1 is given approximately by

$$I_1/I = \lambda/16\pi^2 d$$

where $d = R_1 - R_0$, and may be neglected in comparison with $2R_1$. Values of I_1/I are given in Table 1.

Table 1

	d/λ	10	10^2	10^3	10^4	10^5	
$10\log_{10}$	(I_1/I)	−32	−42	−52	−62	−72	dB

This variation in intensity is similar to that found in the shadow zone of a straight edge diffracting a cylindrical wave front.

So far we have assumed that the reflection coefficient falls abruptly to zero at the head of the trail. The electron line density in trails giving an observable head echo must be greatly in excess of 10^{12} cm^{-1}. For such trails diffusion will cause the reflection coefficient to increase with distance along the trail from the head, at a rate depending upon the velocity of the meteor and the ambipolar diffusion coefficient [6]. This effect will tend to reduce the head echo. However, taking account of the heat liberated in the column by the passage of the meteor, the initial rate of radial diffusion may be considerably greater than the average rate over the lifetime of the body echo, and simple considerations show that the consequent reduction of the head echo will certainly be less than 5 dB.

The above results may be compared with the observations of Millman and McKinley. We will take as an example the Lyrid Meteor II, of reference [3], which exhibits a pronounced receding head echo. The radiant elevation is given as 6°, and the range difference between the body echo and the point at which the head echo vanishes is 7000λ. Hence, from Table 1, $10\log_{10}(I_1/I) \doteqdot -60$ dB. Using the theory of Kaiser and Closs [6] we find that the electron line density required to produce a body echo 60 dB above the noise level is of the order of 10^{16} cm^{-1}, and so we may take this as the line density at the head of the trail at the instant when the head echo becomes undetectable. However, this value of line density may not be that appropriate to the part of the trail returning the body echo, because the ionization produced by a meteor varies with height. It is probable

that the meteor was producing its maximum ionization near the point where the head echo vanished. This point was 13 km lower than that giving the body echo, and by using Herlofson's expression [7], which gives the variation of meteor ionization with pressure, it can be calculated that the line density appropriate to the body echo must have been of the order of 10^{15} cm^{-1}. This value is consistent with the 8 sec duration of the body echo if the ambipolar diffusion coefficient is about 10^6 cm^2 sec^{-1} [6], i.e. if the body echo originated at a height of about 110 km. Thus the height of maximum ionization was probably about 97 km, a reasonable figure for such meteors.

Similar results are obtained when the theory is applied to other meteor echoes described in references [1], [2], [3]. In a very few cases the main echo is not observed. This would occur if the line density at point O were of the order of, or less than, 10^{12} cm^{-1} (in which case O would be near either the beginning or the end of the trail). The resulting echo duration would then be considerably less than one second and might not be resolved on the range-time display.

The rate at which meteors exhibiting head echoes are observed will depend critically on the sensitivity of the radio equipment, because, for very dense trails, the echo amplitude increases only as the fourth root of the line density [6]. For instance, if the equipment sensitivity were some 10 dB greater than that used by Millman and McKinley, a head echo such as that described would be observed from a trail producing only 10^{14} electrons per cm.

The results of the theory proposed here may be summarized thus: the intensity of the echo from the head of a uniform meteor trail relative to the body echo should be proportional to the radio wavelength λ, and inversely proportional to d, the difference in range between the two echoes. The relative intensity will depend also upon the variation of electron line density along the trail. A corollary is that additional, stationary, echoes may be produced by bursts of ionization associated with the visual flares observed along the trails of very bright meteors.

VOL.
4
1953

References

[1] McKinley, D. W. R. and Millman, P. M.; Proc. I.R.E. 1949 **37** 364. [2] McKinley, D. W. R.; J. App. Phys. 1951 **22** 202. [3] Millman, P. M. and McKinley, D. W. R.; J. Roy. Ast. Soc., Canada, 1948 **42** 121. [4] Hey, J. S., Parsons, S. J. and Stewart, G. S.; Mon. Not. Roy Ast. Soc. 1947 **107** 176. [5] Lovell, A. C. B. and Clegg, J. A.; Proc. Phys. Soc. 1948 **60** 491. [6] Kaiser, T. R. and Closs, R. L.; Phil. Mag. 1952 **43** 1. [7] Herlofson, N.; Phys. Soc. Rep. Prog. Phys. 1948 **11** 444.

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 5 to 9. Pergamon Press Ltd., London

Rotational temperatures of auroral nitrogen bands

W. PETRIE

Radio Physics Laboratory, Defence Research Board, Ottawa

(*Received* 20 *October* 1952)

ABSTRACT

Auroral spectrograms of dispersion 42–28 A/mm have been analyzed with a recording microphotometer Rotational temperatures were determined from the profiles of Vegard-Kaplan and Negative Group nitrogen bands, and found to be in the range 450–750°K. It now appears that spectroscopic, rocket, and radio results, are in reasonable agreement.

INTRODUCTION

Although the various determinations of the temperature of the atmosphere up to the auroral layer are in fair agreement, this is not the case in the layer itself. The particle densities (HAVENS, 1952) in the height range 100–160 km as determined from rocket flights, indicate temperatures of 240°K and 570°K at these levels. Radio measurements (APPLETON, 1937) point to still higher temperatures in the F_1 and F_2 layers. In contrast, there are the extensive measurements of rotational temperatures by VEGARD and colleagues which have been summarized in HARANG'S recent book (1951). The mean rotational temperature of the N.G. N_2^+ bands appears to be of the order of 230°K. A slightly lower value has been found by VEGARD and KVIFTE (1951) from the analysis of a recently secured higher dispersion spectrogram. Spectra of sunlit aurora and of the summits of auroral rays appear to yield temperatures of the same order (HARANG, 1951). The discrepancy between these latter results and the rocket and radio measurements is a serious one, and must be removed or explained in a satisfactory manner. This paper discusses some recent determinations of rotational temperatures from moderate-dispersion auroral spectrograms.

THE OBSERVATIONAL MATERIAL

The spectrograms on which this discussion is based and the instrument used to secure them, have been described elsewhere (PETRIE and SMALL, 1952). The plates selected for a study of rotational structure were obtained from the second and third orders of the grating, the dispersions in these orders being 42 and 28 A/mm respectively. From these spectrograms the rotational temperatures of the Negative Group, Vegard-Kaplan, and First Positive Group Nitrogen bands may be estimated.

(a) *Vegard-Kaplan Bands*

This band system arises from the transitions $^3\Sigma_u^+ - {}^1\Sigma_g^+$ and being forbidden, should give a reliable value for the temperature of the region from which the emission arises. Regardless of how the molecules enter the excited electronic state, the lifetime of this state is sufficiently long to ensure that rotational equilibrium is established.

One method of estimating the temperature is as follows. The intensity of a rotational line may be written as,

$$I = c\nu^4 S e^{-B'K'(K'+1)hc/kT} = c\nu^4 S e^{-\frac{1{\cdot}44B'}{T}K'(K'+1)} \tag{1}$$

The frequency factor ν^4 varies little within a band and may be neglected. S is the "line strength" factor, B' the rotational constant of the upper state, K' the rotational quantum number of the upper state, and T the temperature. In the case of the Vegard-Kaplan bands the P and ${}^PQ$ branches overlap, and the S values of the two branches add together. According to SCHLAPP (1937) the value of S is $(K'' + 3/2)$, K'' referring to the lower ${}^1\Sigma$ state. Then,

$$I = c(K'' + \tfrac{3}{2})e^{-\frac{1{\cdot}44}{T}B'K'(K'+1)} \tag{2}$$

Since $K'' = (K' + 1)$ for the P and ${}^PQ$ branches,

$$I = c(K' + \tfrac{5}{2})e^{-\frac{1{\cdot}44}{T}B'K'(K'+1)} \tag{3}$$

At the maximum of the band $\dfrac{dI}{dK'} = 0$. Forming this derivative from (3) and solving for T we find,

$$T = 1{\cdot}44B'(2K' + 1)\,(K' + 5/2) \tag{4}$$

The lines of the P branch may be represented by the formula

$$(\nu - \nu_0) = -(B' + B'')K'' + (B' - B'')K''^2$$

Substituting the constants of the (1,11) Vegard-Kaplan band,

$$(\nu - \nu_0) = 3{\cdot}23K'' - 0{\cdot}37K''^2 \tag{5}$$

The distance from the zero gap ν_0 to the maximum of the P branch of the (1,11) band appears to be of the order of 14A or 103 cm^{-1}. From (5) the corresponding K'' value is 13; hence $K' = 12$. Then from (4) $T \sim 750$°K. The mean temperature derived from five Vegard-Kaplan bands in the wavelength range 3400–3800A is 850°K. These results have already been reported (PETRIE, 1952), and refer to the spectrum of an intense auroral display which occurred on February 23, 1952.

The figures quoted here are preliminary and it is very desirable that denser, higher dispersion spectra of the Vegard-Kaplan bands be secured. Nevertheless, the plates examined show the rotational structure clearly, and the profiles of the bands cannot be reconciled with temperatures of the order of 230°K. It is worth noting that the spectrograph was pointed near the summit of the visible aurora during the exposure, and was likely receiving radiation from heights considerably above 100 km. Furthermore, the greater than normal intensity of the Vegard-Kaplan bands in this spectrogram suggests that the radiation originated at higher levels. Fig. 1 shows the (1,11) and (2,12) bands.

(b) *First Negative Bands*

This system arises from the transitions ${}^2\Sigma_u^+ - {}^2\Sigma_g^+$. With moderate dispersion

VOL. 4 1953

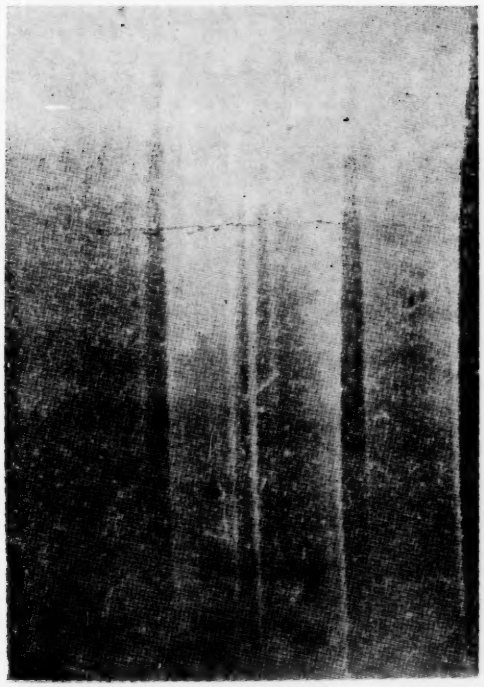


Fig. 1. (1,11) and (2,12) Vegard-Kaplan bands.

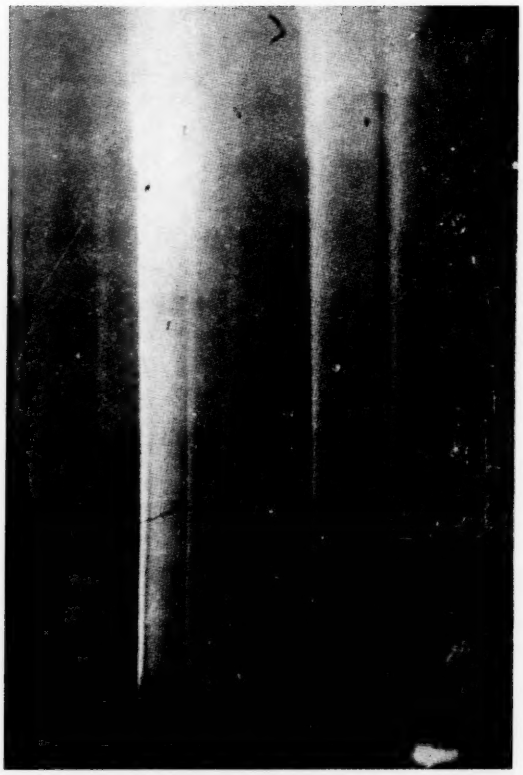


Fig. 2. (0,0) Negative Group band.

VOL.
4
1953

each band shows a single P and R branch; overlapping Q branches of low intensity may be neglected. The R branch extends to shorter wavelengths and the rotational structure may be seen. The spectrograms discussed here have sufficiently high resolution to show the alternate weak lines of the R branch.

According to MULLIKEN (1931) the S factor of the R branch lines is $S = \frac{(J'' + 1)^2 - \frac{1}{4}}{(J'' + 1)}$ and since $J' = (J'' + 1)$ for the R branch, $S = \frac{J'^2 - \frac{1}{4}}{J'}$. Each line is an unresolved doublet with J' values $J' = (K' + \frac{1}{2})$ and $J' = (K' - \frac{1}{2})$. Hence $S_1 = \frac{K'^2 + K'}{K' + \frac{1}{2}}$, $S_2 = \frac{K'^2 - K'}{K' - \frac{1}{2}}$ and $(S_1 + S_2) = \frac{2K'^3 - K'}{K'^2 - \frac{1}{4}}$. For our purpose a satisfactory approximation is to write $S = 2K'$. Then,

$$I = c\nu^4 2K' e^{-\frac{1{\cdot}44B'}{T}K'(K'+1)} \tag{6}$$

Forming the derivative $\frac{dI}{dK'}$, and equating to zero we find,

$$T = K'(2K' + 1)\, 1{\cdot}44B' \tag{7}$$

The formula for the R branch is,

$$(\nu - \nu_0) = 2B' + (3B' - B'')K'' + (B' - B'')K''^2 \tag{8}$$

Substituting the values of the rotational constants gives the following expressions for the (0,0) and (0,1) bands.

(0,0) $$(\nu - \nu_0) = 4{\cdot}2 + 4{\cdot}3K'' + 0{\cdot}15K''^2 \tag{9}$$

(0,1) $$(\nu - \nu_0) = 4{\cdot}2 + 4{\cdot}4K'' + 0{\cdot}18K''^2 \tag{10}$$

A few of the sharpest plates were selected and analyzed with a Jarrell-Ash recording microphotometer. A picture and a trace of the (0,0) band at 3914A are shown in Figs. 2 and 3. The resolution is such that the intensities of a number of the individual lines may be determined, and it appears that the position of the intensity maximum is $\sim$59 cm^{-1} from the zero gap. From equations (9) and (7), $K'' = 10$, $K' = 11$, and $T \sim 750$°K. Even if the maximum is displaced two lines towards the zero gap, the wave-number difference between this line and the zero gap is 47 cm^{-1}, and $T \sim 500$°K. The temperature range determined from the (0,1) band at 4278A was 500–600°K.

From (6) it can be seen that,

$$\log \frac{I}{K'} = \log c - \frac{1{\cdot}44}{T} B'K'(K' + 1) \tag{11}$$

Hence, by plotting the factors $\log \frac{I}{K'}$ and $K'(K' + 1)$, the temperature may be determined from the slope. The I factors were obtained by converting the profiles of the lines on the microphotometer tracings from a logarithmic to a direct intensity scale, and them measuring the areas with a planimeter. On the other hand, since the individual rotational lines are very narrow it is satisfactory to use the central

intensity of each line as a measure of its total intensity. The spectra were secured with an instrument which used a step-slit, and each spectral feature acts as a source from which the "characteristic curve" of the emulsion may be determined. Plots were made using both central intensities and total intensities for the I values, and the derived temperatures were in the range 450°-700°K.

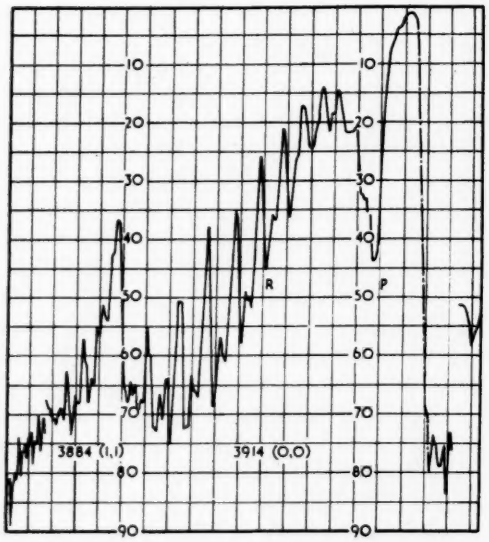


Fig. 3. (0,0) Negative Group band

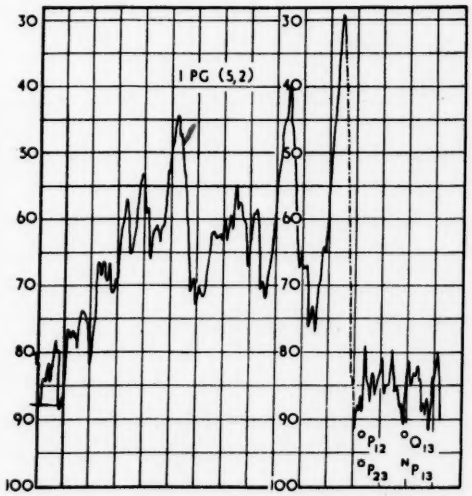


Fig. 5. (5,2) First Positive Group band

The question now arises, why is there a serious discrepancy between these results and those of VEGARD and his collaborators? The method used is quite straightforward, and the error involved in the temperature determination is much smaller than the differences between the two sets of results. As far as the writer can judge from VEGARD's recent paper (1951), the latter has picked out the portion of the band corresponding to the smallest microphotometer deflection, and has called the line at this point the one with maximum intensity. But underlying the lines in this portion of the band there is a continuous spectrum due to convergence of the rotational lines and possibly due to the presence of the (3,6) Second Positive Band (PETRIE and SMALL, 1952). If the background is caused by a foreign band, then it is the blackening of the line relative to the background, rather than the total blackening which determines the line intensity. If the background is caused by the overlapping of rotational lines, the relationship between true and measured intensity is complex, and the interpretation of the profile of this region of the band must be treated with caution.* Hence the relative intensities of the lines well away from the head should give the best indication of the temperature. The spectra analyzed by the writer are of relatively high resolution, the dispersion being in the range

* The writer wishes to thank Dr. A. H. DOUGLAS of the National Research Council of Canada for discussions on this point.

VOL. 4 195

OL.
4
953

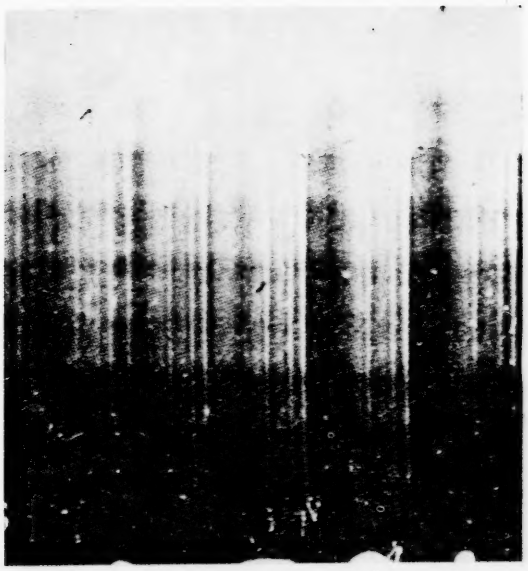


Fig. 4. First Positive Group bands

VOL.
4
1953

42–28A/mm, and it was possible to measure the intensities of individual lines. This has not been done in the past. It is worthwhile noting that the profile of the (0,1) band at 4278A appears similar to the illustrated profile (HERZBERG, 1950) at a temperature of 673°K.

(c) *First Positive Bands*

These bands, from the transitions ${}^3\pi_g - {}^3\Sigma_u^+$ are very complex, and are poorly resolved in the spectra discussed here. As can be seen from Figs. 4 and 5 each band consists of three sharp regions and three diffuse regions. Each of these regions includes several branches. Since the line structure of these bands cannot be resolved, only a rough estimate may be made of the rotational temperature.

The "line strength" factors depend upon whether the coupling approaches HUND'S cases (a) or (b). NAUDÉ (1932) has shown that the coupling is according to case (a) for small rotational quantum numbers, and approaches case (b) for larger quantum numbers. "Line strength" factors have been given by NOLAN and JENKINS (1936) for both cases (a) and (b). The simplest part of one of these bands is the diffuse region to the long wavelength side of the first strong head and this region consists of the branches ${}^0P_{12}$, ${}^0Q_{13}$, ${}^0P_{23}$, and ${}^NP_{13}$, the latter branch being weak compared with the other three. If we substitute the "line strength" factor in the intensity equation and form the derivative $\frac{dI}{dK'} = 0$, then it appears that for each of the first three branches given above an approximate expression is $T = J'(2J' + 1)\ 1{\cdot}44B'$. From the apparent position of the maximum of the region referred to, the writer judges that the temperature is 350°K. The accuracy of this figure is not comparable to the temperature values derived from the other band systems.

CONCLUDING REMARKS

In view of the fact that the results quoted here contradict earlier spectroscopic investigations, it is important that additional observations be made with higher resolution equipment. Several spectrographs now under construction will be adequate for a careful study of the rotational structure of auroral bands, and a definite decision regarding spectroscopic temperatures should soon be reached.

REFERENCES

APPLETON, E. V.	1937	Proc. Royal Soc. London A. **162**, 451
HARANG, L.	1951	The Aurorae, Wiley and Sons
HAVENS, R. J., KOLL, R. T. and La GOW, H. E.	1952	Jour. Geophy, Res. **57** 59
HERZBERG, G.	1950	Spectra of Diatomic Molecules, 2nd Edition. D. VAN NOSTRAND, page 50.
MULLIKEN, R. S.	1931	Rev. Mod. Physics, **3** 87
NAUDÉ, S. M.	1932	Proc. Roy. Soc. London **136**, 114
NOLAN, P. and JENKINS, F. A.	1936	Phys. Rev. **50**, 943
PETRIE, W.	1952	Phys. Rev. **86**, 790
PETRIE, W. and SMALL, K.	1952	Astrophys. J; **116**, 433
SCHLAPP, R.	1937	Phys. Rev. **51**, 342
VEGARD, L. and KVIFTE, G.	1951	Geofys. Publikasjoner, 18 (No. 3)

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 10 to 27. Pergamon Press Ltd., London

An interpretation of vertical incidence equivalent height versus time recordings on 150 Kc/s*

RUNE LINDQUIST†

Ionosphere Research Laboratory, The Pennsylvania State College, State College, Pennsylvania

(*Received* 21 *November* 1952)

ABSTRACT

Results of virtual height versus time vertical incidence pulse recordings, obtained on 150 Kc/s, are presented and discussed. Monthly median values of the reflection heights are shown in a series of graphs. The coupling echo, predicted by current wave theory, is definitely shown to exist. The results of measurements during undisturbed and disturbed days are discussed. It is concluded that one form of echo regularly noticed during magnetically disturbed nights must be due to one type of sporadic *E*. Recorded groups and phase heights are compared and the differences checked against those predicted theoretically. Finally, the results are given of a preliminary investigation of the effects of solar flares.

INTRODUCTION

Low frequency vertical incidence pulse soundings of the ionosphere have been underway at this Laboratory since 1949. The part of the work dealt with in this paper is the recording and interpretation of virtual height versus time. The experimental work considered covers the period January 1950—April 1952.

Due to various reasons, we have not been able to obtain a complete coverage of data for this period. As far as the normal behaviour of the lower *E*-layer is concerned, this paper, therefore, will discuss results obtained during the following months: January 1950; February, March, August–December 1951; and January–March 1952. Most of the recordings obtained during the entire period can, however, be used for investigations of special phenomena; for example, the behaviour of the *E*-region during disturbed conditions, solar flare correlations, sunrise phenomena, etc. The paper also gives a survey of the information hitherto obtained on these questions.

Results of low frequency virtual height soundings using the pulse technique have been reported earlier by HELLIWELL (1950, 1952) and BLAIR, BROWN and WATTS (1950, 1951, 1952). Some of their results will be compared with those obtained in this paper.

GENERAL DISCUSSION CONCERNING THE RECORDS AND THEIR INTERPRETATION

For this recording programme we have used the conventional method of transmitting a short pulse of radio frequency. The direct or ground pulse and the pulses reflected from the ionosphere are picked up at the receiving site, detected, applied as intensity modulation on a cathode ray tube, and photographed on moving film in a conventional manner.

* The research reported in this paper has been supported by the Geophysics Research Division of the Air Force Cambridge Research Centre under Contract AF19(122)-44 and by the Department of State under Contract SCC-2680.

† On leave from the Chalmers University of Technology, Gothenburg, Sweden.

VOL. 4 1953

This system has certain limitations when applied to a low frequency carrier frequency; namely, the problem of obtaining a sufficiently steep leading edge and short duration of the transmitted pulse. By using some compromises a satisfactory solution has been found, the resulting transmitted pulse, approximately Gaussian, having a risetime corresponding to between five and ten kilometres virtual height and a total duration of about 350μ sec.

Our main interest in the present work is to study the fine structure of the lower region of the ionosphere; that is, we should be able to measure the recorded virtual heights of reflection within a few kilometres. When recording a transmitted pulse of the above-mentioned shape, we may choose between one of two methods. One method is to preserve the shape of the ground pulse and the echoes throughout the receiver. The virtual heights can then be found very accurately by scaling, for example, from the peak of the ground pulse to the corresponding peak of the echo. This method has been discussed in a recent paper by WHALE (1951). The other method of accomplishing the same degree of precision is to utilize a limiting amplifier in the video stages of the receiver. The records can then be scaled from the leading edge of the ground pulse to the leading edge of the echo. For the sake of simplicity we have chosen to use the latter system. The few disadvantages, for example the tendency of giving too high readings for very weak echoes, can easily be overcome by careful scaling.

It is also necessary to include a differentiating circuit in the video amplifier. This makes it possible to, within certain limits, record the closely spaced split echoes which appear quite frequently on the night-time records, as discussed later in this paper.

A series of sample records illustrating most of the phenomena observed appear in Figs. 1–3. These are direct recordings obtained by the movement of a suitable photographic paper past an image of the recording oscilloscope trace. Time increases from left to right for each record and is in local mean solar hours as indicated under each interruption of the record which is used as a time marker. Equivalent height increases vertically, the bottom solid trace being the direct or ground pulse while solid traces above this are ionospheric echoes. The thin, uninterrupted, horizontal lines are height marker traces. These are spaced 10 kilometres virtual height apart. In scaling these records it is necessary to add 4 Km to account for the transmitter-receiver separation of this amount. The interpretation of these records is discussed in a following section.

As far as routine scaling is concerned, the same principles as those internationally accepted for scaling equivalent height versus frequency records have been used. These are summarized in, for example, the C.R.P.L. publication, "Ionospheric Data, Series F"*.

In order to obtain average monthly values for the heights of the *E*-layer reflections, we have used the system of computing median values. We have considered the part of the *E*-layer we are interested in, namely, the bottom, as regular in its diurnal behaviour as the higher parts of the layer. However, as will be discussed later in this paper, the night-time *E*-layer may be much more irregular

* "Ionospheric Data, Series F," issued by C.R.P.L., Nat. Bur. of Stand., Washington D.C., January and February, 1952.

in its behaviour than the daytime layer, especially during magnetically disturbed conditions.

THE DIURNAL VARIATIONS OF THE *E*-LAYER REFLECTIONS UNDER UNDISTURBED CONDITIONS

Before discussing the experimental results regarding the diurnal variations of the *E*-layer reflections at 150 Kc/s, let us briefly summarize certain theoretical work which has appeared in earlier publications from this Laboratory (GIBBONS and NERTNEY, 1952a, b). These consider the wave equations for the two magneto-ionic components, π_1 and π_2,

$$\begin{aligned} \pi_1'' + (k_0^2\varepsilon_1 + M^2)\pi_1 &= -\pi_2 M' - 2\pi_2' M \\ \pi_2'' + (k_0^2\varepsilon_2 + M^2)\pi_2 &= -\pi_1 M' - 2\pi_1' M \end{aligned} \tag{1}$$

where $M = \dfrac{u'}{1 - u^2}$ is the coupling factor and u is the polarization of the wave as discussed in detail in the references. A set of certain models of the electron distribution and the collisional frequency in the height range of interest, in the *E*-layer, are assumed. It is shown that, for the frequency of interest, there should result in two returned echoes, under suitable conditions, originating in the vicinity of the 300 and 3000 electrons/cm^3 points in the *E*-layer. The signal returned from the 300 el/cm^3 region is shown to be mainly due to the coupling factor terms in Eq. (1). We will call this region the coupling region. The high collisional frequency at this level prevents the existence of reflection conditions during daytime hours. During the night, the height of the 300 electrons/cm^3 level increases, taking this region into a lower collision level. As shown in the references, this increases the coupling effect with consequent reflection from this level.

The other electron density of interest is the region where $N = 3000$ electrons/cm^3. Here the conditions for reflection of one of the magneto-ionic components is fulfilled. This region will, therefore, be referred to as the main reflection region. We will, in general, refer to the signals from these two levels as coupling echo and main reflection echo.

It should be emphasized that this theoretical work, since it is based on the use of continuous waves, does not account for the dispersive properties of the medium. An extension of a portion of this work to include the dispersion of the medium appears elsewhere, DAVIDS (1952). This will be referred to later in this paper.

Returning now to Figs. 1–3, an inspection of the group height records shows that the occurrence of split echoes is very common during certain hours of the day. See, for example, Fig. 2B. From the beginning of the record up to about 2330 LMST, we can only distinguish one single first and second hop echo. From 2345–0215 we can very definitely observe two first hop echoes, the centre traces. The separation between the echoes and their respective amplitudes, estimated from the width of the traces, are varying somewhat with time. At 0215, the upper of the split hop echoes decreases in height by about ten kilometres and the lower echo disappears. This condition lasts up to about 0340, when the echoes once more separate and remain separated until sunrise is reached at about 0640. This

VOL. 4 195

OL.
4
953

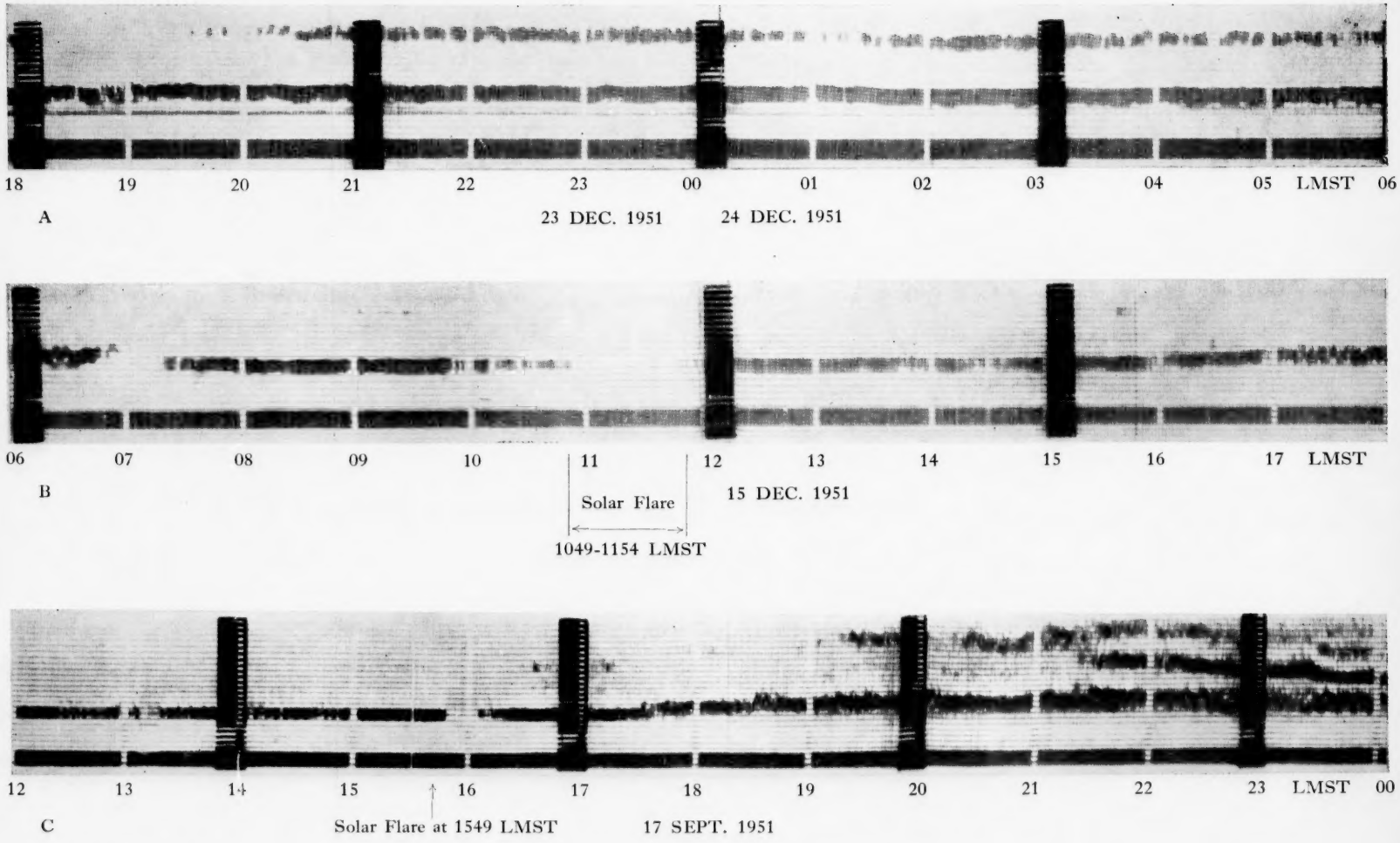
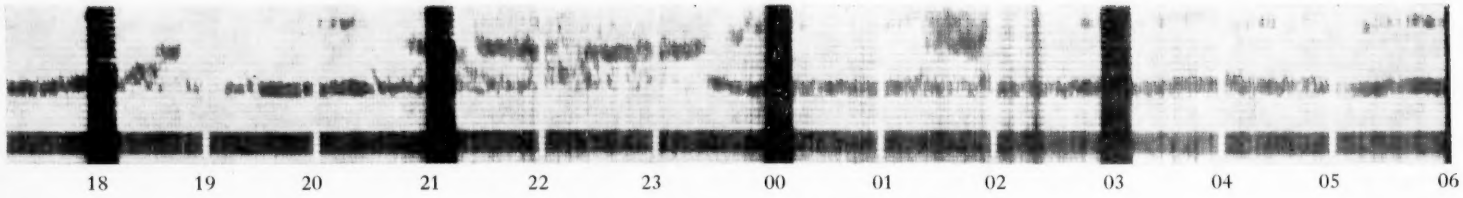
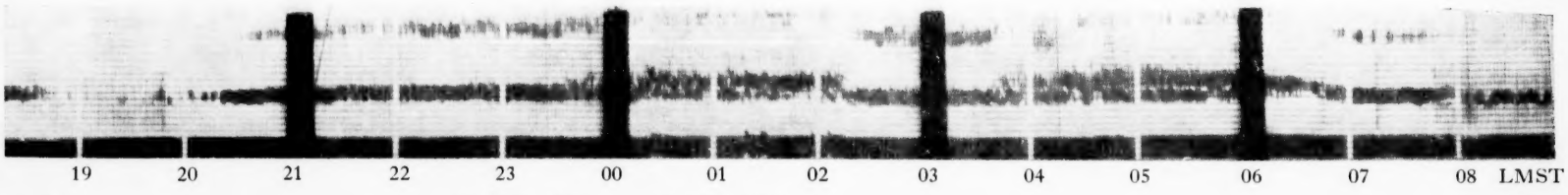


Fig. 1. 150 Kc/s vertical incidence equivalent height vs. time records





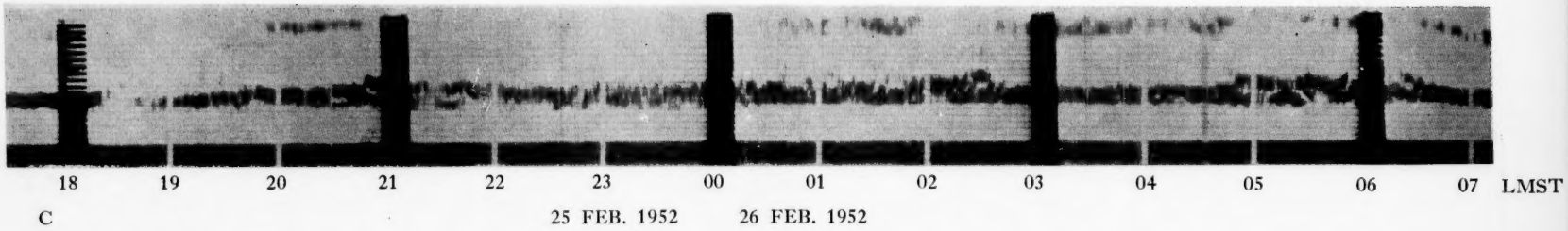


Fig. 2. 150 Kc/s vertical incidence equivalent height vs. time records

VOL.
4
195

latter part of the record, an hour or so before sunrise up till sunrise, is very typical for most nights. The two types of echoes just described are believed to be what we earlier have referred to as main reflection and coupling echo; the latter being, of course, at the lower height under conditions of splitting.

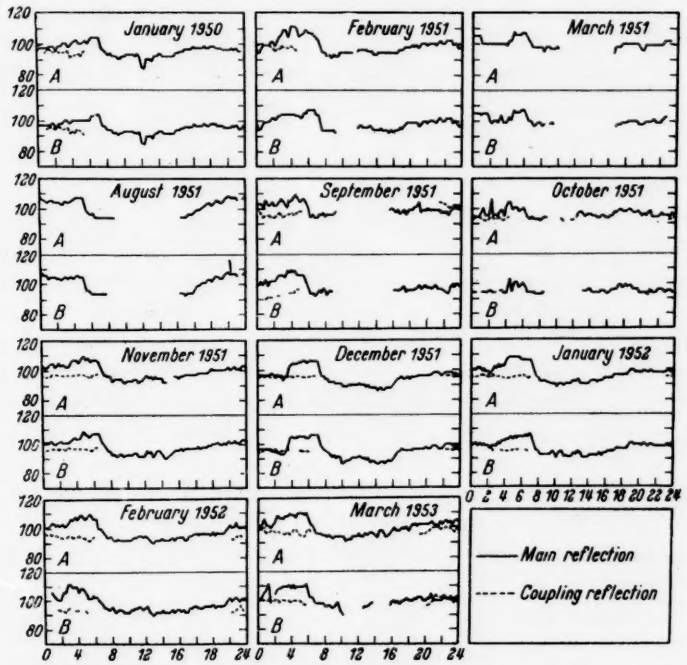


Fig. 4. Monthly median 150 Kc/s vertical incidence equivalent heights vs. diurnal time.
A. Monthly median values for all available data.
B. Monthly median values for days with no significant magnetic disturbance (half-day K-index of Cheltenham, Md. < 4).

In Fig. 1A is depicted another type of night-time record. This record characterizes a, magnetically, very quiet night. We observe, mainly, one single, continuous first hop echo throughout the whole night. There is very little indication of height changes in the layer and split echoes are difficult to detect. However, there are some split echoes present during the later part of the night. The reason why these cannot be recorded as two separate echoes is, as mentioned earlier in this paper, the width of the ground pulse. When the main reflection and coupling levels are separated by less than about ten kilometres, it is very difficult to record them as two separate echoes. The coupling echo will only manifest itself as a downward extension of the main reflection echo. But as, in general, the coupling echo shows more variations in amplitude than the main reflection, this provides a means for detecting both components during the early morning hours.

Finally, in Fig. 1B is illustrated a daytime record. This record shows a single reflection throughout the day with no sign whatsoever of a coupling echo. The

interruption of the echo during the interval 1049–1155 LMST is, as explained in a later section of this paper, caused by a solar flare. Normally, at this time of the year, we would expect to observe an echo continuously throughout the whole day.

In the routine scaling of the records, we have been mainly interested in establishing a definite proof for the existence of the two above-mentioned reflections from the *E*-layer and also to find their diurnal and seasonal variations. The monthly median heights of reflection for both echoes have been computed and are shown in Fig. 4. These graphs definitely establish the existence of the two reflecting regions during the early morning hours. The theory also predicts the greatest separation of the echoes, and the strongest amplitude of the coupling echo, at low critical frequencies and increased layer heights, a condition which is most likely to exist in the early morning hours.

A closer inspection of the diurnal heights of the coupling level reveals that, on some months, the *median* virtual heights of the coupling echoes are greater than the corresponding *median* heights for the main reflection echoes. This condition indicates that our method of determining median heights must, on occasion, be in error. As a method of tracking down this error, all magnetically disturbed days during these months were selected and excluded from the computation. The computed median values of the reflection heights using only undisturbed days are also plotted on Fig. 4. However, as the Figure shows, we still have occasional cases where the median height of the coupling echo exceeds the height of the main reflection echo. The reason for this must, therefore, be sought in the number of the individual values used for computing a median value. In this work we have considered five or more values as satisfactory for computing a median value. This is definitely justified for the daytime echoes. But the abnormal behaviour of some of the night-time median plots indicates that the *E*-layer at that time is very irregular. It would, therefore, be preferable to consider a median value based on less than five values as unacceptable, as also has been done in this paper. But a value based on five-nine values should be considered as questionable, instead of certain as we have done. However, the error introduced by accepting less than ten values as giving a good median value is, in general, probably not too great.

In Table 1 below we have listed some typical heights for different quantities. It is very interesting to note, for example, that the median virtual height of the coupling region is almost constant throughout the year, ranging between 93 and 96 kilometres. This height will very closely correspond to the bottom of the night-time *E*-layer. A comparison with the smoothed noon values of reflection heights gives about the same heights; the noon values being slightly lower, the difference varying from zero to seven kilometres.

We have considered two different height values for noon (A and E). The reason for this is quite obvious if, especially, the latter, and more reliable, data of Fig. 4 are examined. Inspection of the various monthly plots shows that the main reflection first decreases in height from sunrise up to an hour or so before noon. Thereafter, centred near noon, we can see, on most of the monthly plots, a very definite tendency of increasing height. This increase reaches a maximum somewhat after noon, after which the reflection height starts to decrease. However, the decreasing part of the curve is not so pronounced; it merely merges into

VOL. 4 1953

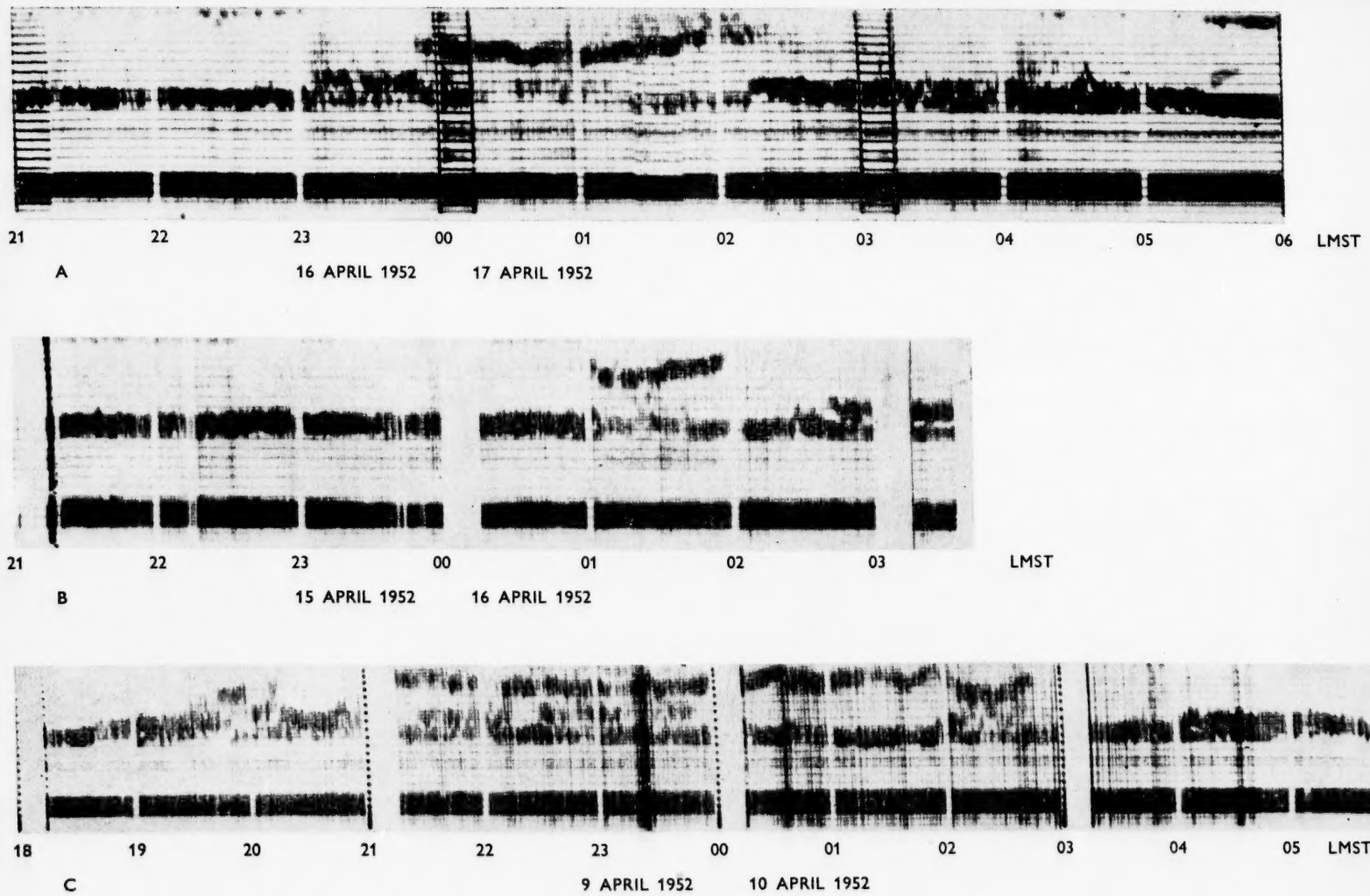


Fig. 3. 150 Kc/s vertical incidence equivalent height vs. time records

VOL.
4
1953

Table 1. *Various heights as deduced from the monthly median plots. All heights are given in kilometres. Numbers in brackets are considered doubtful because of lack of sufficient data.*

Height (km)	*Jan.* 1950 *Jan.* 1952	*Feb.* 1951 *Feb.* 1952	*Mar.* 1951 *Mar.* 1952	*Aug.* 1951	*Sept.* 1951	*Oct.* 1951	*Nov.* 1951	*Dec.* 1951
A. Smoothed Noon Value	90 90	92 91	(95) 92	(94)	(94)	93	92	88
B. Night Peak Main Reflection	104 108	111 110	107 110	107	109	106	109	106
C. Smoothed Night Coupling	93 94	96 93	— 96	—	95	93	96	95
D. Smoothed Night Main Reflection (Max. Value)	102 106	108 107	106 110	(106)	105	101	107	106
E. Maximum Value Noon ±1 Hour	93 93	96 95	— 96	—	94	97	97	91
B–A	14 18	19 19	(12) 18	(13)	(15)	13	17	18
C–A	3 4	4 2	— 4	—	(1)	0	4	7
D–A	12 16	16 17	(11) 18	(12)	(11)	8	15	18
E–A	3 3	4 4	— 4	—	0	4	5	3

the expected afternoon increase of the reflection height. Since the possible errors involved in the determination of the reflection height, as discussed earlier in the general discussion, have positively been removed in the scaling procedure, we are forced to accept this increase in noon height as real. The most likely explanation is to attribute the increase to an extra retardation in the *D*-region. We have, therefore, listed in Table 1 both the maximum median height recorded near noon, E, and the smoothed median height at noon, A. The latter height was determined by plotting the best smooth curve through the daytime data, excluding the part near noon. This curve should, therefore, correspond very closely to the changes in the *E*-layer. We can see from Table 1, E–A, that the extra group retardation, possibly introduced by the *D*-layer, is of the order of three to five kilometres. It is indeed very unfortunate that we are unable to obtain values for the period April—July due to the lack of system gain. This information should be very valuable, especially for establishing seasonal trends.

The very abrupt decrease in reflection height near sunrise is characteristic for all months. This problem has been studied in detail by the author elsewhere

(1952). No further discussion of this phenomenon should, therefore, be necessary at this time.

A comparison of the median curves obtained for the same months in different years shows a very close correspondence. The median values for January 1950 and January 1952, for example, are identical within a few kilometres. This, perhaps, can be taken as an indication that solar activity has less influence on the ionization near the bottom than at the maximum of the *E*-layer.

As mentioned earlier, it has been shown (DAVIDS, 1952), that the effect of the dispersion in the medium for our pulse signals would be, under certain

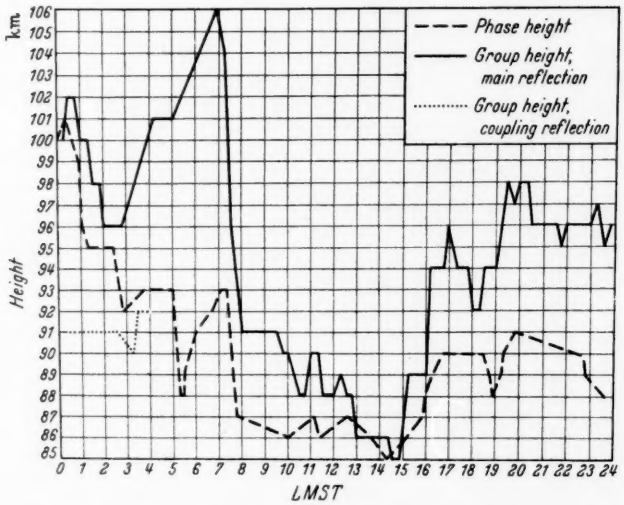


Fig. 5. Comparison between the diurnal variations in group and phase heights on 150 Kc/s for a typical day.

conditions, to increase the time delay of the main reflection echo. That is, increase the group height compared to the phase height. As indicated in the reference, the maximum time delay is to be expected at low *E*-layer critical frequencies. Efforts have been made to verify this theory by comparing group and phase height records. The phase height records have been obtained by R. E. JONES of this Laboratory. His recording programme is to be described elsewhere at a later date.

In Fig. 5 is shown one of the representative diurnal phase height records together with the group heights for the same day, 23 December 1951. As the phase height recordings only give *changes* in phase height, and not the absolute phase height, we must match the two records at one point. This was done at the point during daytime where both records showed the lowest height. We, consequently, assume negligible difference in phase and group height at that time, a procedure which appears to be valid in view of DAVID's paper. It can be seen from this Figure how changes in group and phase height correspond throughout the day, the relative difference in the two heights being of the order of three to four kilometres

VOL. 4 1953

immediately after sunset when the corresponding E-layer critical frequency probably is fairly low. This difference decreases at a fairly fast rate during the hours immediately following sunrise, the difference in height being of the order of one kilometre near noon. This behaviour checks very well with the predicted values given in the paper by DAVIDS already referred to. Also, the night-time behaviour gives differences of about the same order of magnitude as those predicted.

The only part of the phase-height records that has to be considered as doubtful in this example is the part from 0245 to 0715. During this time an increase in group-height corresponds to a decrease in phase-height. However, as the group-height record shows, coupling reflections were present immediately preceding this interval. The recorded phase-heights are, therefore, possibly contaminated by the coupling reflections. It is, consequently, impossible to estimate the maximum difference in phase and group height in this interval. It would appear to be of the order of eight kilometres.

THE REFLECTIONS ON 150 KC/S DURING MAGNETICALLY DISTURBED CONDITIONS

It has been found during this recording programme of virtual reflection heights versus time on 150 Kc/s that there is a considerable difference between the behaviour of the E-layer during a disturbed and an undisturbed *night*. This fact is in contrast to the behaviour of the *daytime* layer for the two conditions. Here the only difference during a magnetic storm is an increase in daytime absorption with no significant change in the shape of the $h' - t$ curves when compared with a preceding undisturbed day.

Records obtained during disturbed nights are characterized by a very irregular behaviour of the reflections. Sudden jumps from one reflecting level to another appear and the virtual heights of reflection show comparatively rapid fluctuations. Traces showing the typical behaviour of changing retardation are obtained quite frequently.

A selection of commonly observed records obtained during magnetically disturbed conditions appear in Figs. 2 and 3. The most predominant effect to be observed during these conditions is the appearance of a new reflecting stratum somewhat above the normal E-layer. Fig. 2A depicts a typical example of this new reflecting medium, the reflection lasting from 2044 to 2345 LMST. This Figure also clearly demonstrates the behaviour of the normal E-reflections during the period of the disturbance. These reflections seem to nearly disappear. We will return to this question somewhat later in the paper.

As a first attempt to obtain information on the sporadic reflections above the E-layer, their frequency of appearance was studied. The day was divided into three-hour intervals corresponding to the three-hour divisions of the Greenwich Day. The total duration in hours of the presence of sporadic reflections within every three-hour interval was compiled and tabulated and is herein titled the index of intensity of sporadic reflections. This is the basic information for comparing these echoes with certain other manifestations of magnetic activity. It may be mentioned in this connection that during certain of the months of interest, it has not been possible to obtain this information. The reason is that the height

range of the recording system had been adjusted to a restricted height range, too low to record these reflections. These conditions have, especially, prevailed during November, December 1950 and May—August 1951.

In Fig. 6 is shown a typical example of the three-hour range index of intensity of the sporadic reflections. The index is plotted for every three-hour period during March and April 1952. There is also plotted the three-hour range magnetic K-index

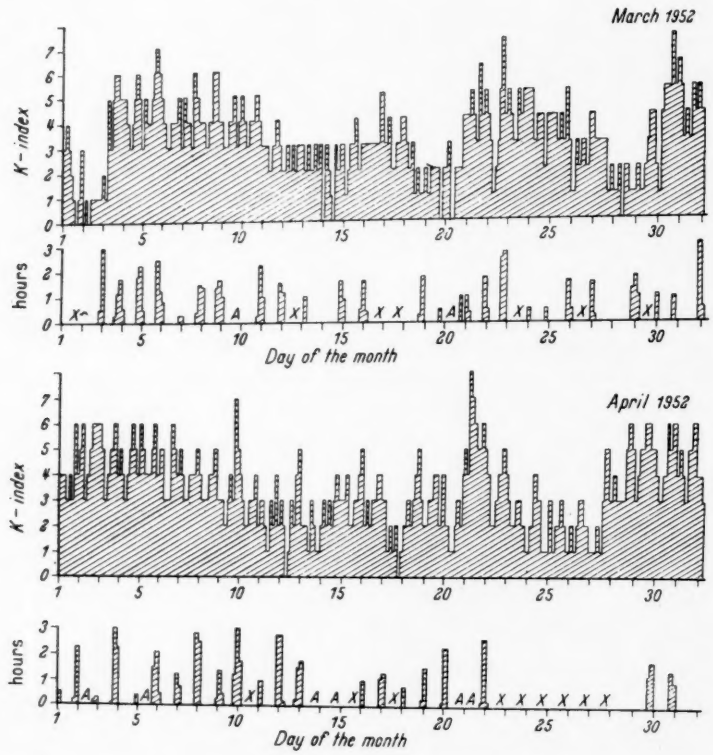


Fig. 6. Correlation between the magnetic K-index and Index of intensity of sporadic reflections above the E-layer for two typical months.

as measured at Cheltenham, Maryland* for these months. Besides the vertical lines indicating the intensity of the sporadic reflections, there appear certain letter designations on the Figure. These are defined as follows:

A = Record unsatisfactory for some reason that particular night.
X = Record satisfactory but positively no sign of sporadic reflections during that particular night.

Diagrams similar to Fig. 6 have been plotted for the period January 1950—April 1952 and all show the same general tendency as Fig. 6. This Figure clearly

* Published for example in "Ionospheric Data, Series F," issued by C.R.P.L., Nat. Bur. of Stand., Washington D.C.

VOL. 4 195

demonstrates the close correlation between the intensity of the sporadic reflections and the magnetic activity both in detailed and in "gross" behaviour.

In order to have a closer check on the dependence of these reflections on the magnetic conditions every individual reflection was rechecked. The total duration in hours of every single continuous reflection was plotted versus the average K-index during the corresponding time. These plots were made on a monthly

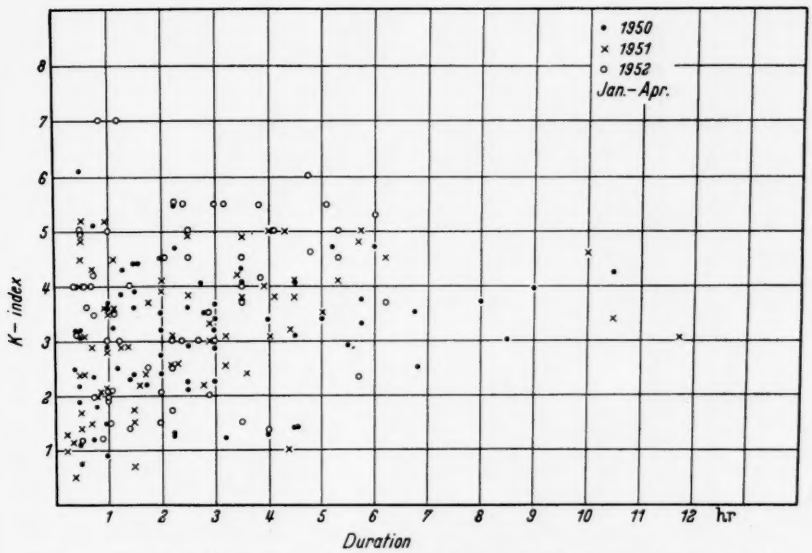


Fig. 7. Mass plot of all data correlating the magnetic K-index with the index of intensity of sporadic reflection.

basis. The general tendency of the individual points on these plots seemed to be to group on or above a straight line with a slope of about 45° and going through the origin of the coordinate system. In Fig. 7, all these individual monthly plots are replotted in the same graph. This Figure very clearly demonstrates the grouping on or above the above-mentioned straight line of the majority of the points. This result appears to indicate that, in order to produce a sporadic echo of a certain duration, the magnitude of the disturbance has to exceed a certain minimum value of the K-index, this minimum value being greater the longer the produced echo lasts.

Finally, the total duration in hours of the sporadic reflections during each month was computed. The result is plotted in Fig. 8. The mean values for the months have been connected to give a picture of the seasonal variation of the sporadic echoes. The general shape of this curve corresponds, for example, to the well-known seasonal distribution of auroral activity, with peaks in March and September.

This plot has also been compared with the indices of auroral activity in North America as published by GARTLEIN *et al.* (1951). However, his published data did not extend beyond June 1950. A satisfactory correlation was obtained for the

period January—May 1950. This is also true when his average curve for the year and our average curve for that and later years are compared. It is especially interesting to note the third maximum in auroral activity in June, obtained by GARTLEIN for the years near the sunspot maximum. The seasonal distribution of the sporadic reflections shows the same maximum in June. However, this point is based on only one year's data, 1950. On the other hand, as this year is closest

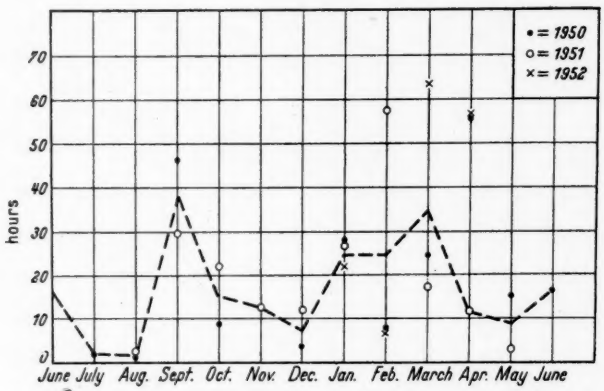


Fig. 8. The seasonal distribution of the appearance of sporadic reflections on 150 Kc/s.

to the sunspot maximum of the three years covered by this investigation, it might well be significant.

Before discussing the possible origin of the sporadic reflections, a few comments will be made regarding the virtual heights of reflection. As can be seen from the sample records in Figs. 2 and 3, these virtual heights usually show typical retardation characteristics. However, in most cases, the echo appears at a fairly great height, decreases in height until a minimum value has been reached, remains for a certain period of time at that height and, finally, disappears with increasing height. We, therefore, have recorded both the hourly heights and the minimum height of the reflection. Of these two quantities, the minimum height is of special interest. The minimum heights recorded during each month in the period January 1951—April 1952 are given in Table 2 below.

Table 2. Minimum virtual heights in kilometres of the sporadic reflections recorded during the period January 1950—April 1952

	Jan.	*Feb.*	*Mar.*	*Apr.*	*May*	*June*	*July*	*Aug.*	*Sept.*	*Oct.*	*Nov.*	*Dec.*
1951	124	124	124	124	143	—	—	133	123	123	137	131
1952	131	143	127	123								

From this Table we can see that the minimum heights appear to have a tendency to fall within the height range 123–131 kilometres. This is also clearly

demonstrated in Fig. 9, where all the individual points during this period have been plotted. There is no pronounced seasonal variation to be noted.

When discussing the possible height of reflection for these sporadic reflections it would be very valuable to have some knowledge of their height versus frequency characteristic. Some tests were, therefore, performed using a manual sweep frequency recorder. The usable frequency range of this recorder is 560 Kc/s—3600 Kc/s. On the number of occasions we have had this manual recorder in

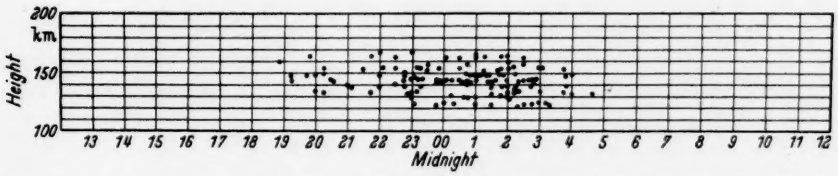


Fig. 9. Mass plot of all data showing the diurnal variation of the minimum heights of the sporadic reflections on 150 Kc/s.

operation, the existence of an echo trace up to about 900 Kc/s has been noted. These reflections have been noted at the same time as we have observed sporadic echoes on 150 Kc/s. The virtual heights of the higher frequency reflections have fallen in the range 135—155 kilometres, with the heights in the individual cases corresponding closely to those noted on 150 Kc/s.

Up to this time it has only been possible on a few occasions to obtain polarization records of the sporadic echoes in view of the height range to which this equipment is normally set. In the cases recorded, mostly manual observations, the polarization has been elliptic (almost circular) with a left handed sense of rotation. Whether this holds for all sporadic reflections is, in view of the scarcity of data, impossible to say, but it is hoped that this question will be settled by a continued check of such records.

What now will all this information on the sporadic reflections tell us? The first conclusion will, of course, be that they manifest themselves only under disturbed conditions. In view of the fact that they appear above the regular *E*-layer we have a choice of two possibilities. Either their origin actually is at this height or they are located below the *E*-layer and the reflections are of oblique origin. Tests performed on antennas with different directive characteristics have indicated that the reflections arrive from the vertical. Another strong argument favouring the statement that the reflections arrive vertically is the agreement between measurements on 150 Kc/s and on 500–1000 Kc/s. We, therefore, assume that these reflections are at vertical incidence.

The reflections found in the same height range at higher frequencies would definitely be classified as a form of sporadic *E*. In view of the close correspondence in height and time between the echoes in these two frequency ranges on all the occasions simultaneous recordings have been made, it seems justified to assume that the sporadic reflections we see on 150 Kc/s also are to be classified as sporadic *E*. In this connection it should be mentioned that sporadic reflections of apparently the same type have been reported earlier by Brown and Watts (1950; 1951;

BLAIR, BROWN and WATTS, 1952) and HELLIWELL (1950, 1952) on low frequencies. Another interesting paper also mentioning reflections of this type is that of GERBER and WERTHMULLER (1947). The last paper discusses reflections on several fixed frequencies in the frequency band 556–1167 Kc/s. The authors did not quote any height measurements, but their published records show heights of the same order of magnitude as those we have obtained. This is a very interesting further check on the theory that our echoes are from one form of sporadic *E*. HELLIWELL (1952) has also pointed out correlations between certain types of low frequency echoes and sporadic *E* on higher frequencies.

As was shown in a preceding paragraph, the maximum night-time *E*-layer ionization usually does not go below 3000 el/cm^3, one of the reflection points for the 150 Kc/s waves. This is the case on quiet nights. On the other hand, during disturbed nights, we have many reasons to believe this is no longer true. See, for example, Fig. 3A. At 2300, the layer appears quite normal with no detectable sign of split echoes indicating we are approaching the critical frequency. However, only ten minutes later the virtual height starts increasing, and at 2315 we have a well-developed split. At about 2348 the upper echo, corresponding to $N = 3000$ el/cm^3, completely disappears. The last part of the trace shows the retardation cusp so typical for an echo just penetrating a layer. At about the same time we note a reflection from a stratum above the *E*-layer. During the first hour this echo shows a decreasing height, while the coupling echo shows increasing height and increasing absorption. At about 0100 the conditions gradually change, the sporadic echo increasing in height and the coupling echo decreasing in height and absorption. Finally, at 0215, the sporadic echo suddenly disappears and, at the same time, both echoes from the *E*-layer show up again.

VOL.
4
1953

Our interpretation of these phenomena is that the maximum electron density of the *E*-layer started to decrease slightly before 2300, passed through a value of 3000 el/cm^3 at 2348, continued decreasing and reached its minimum value at 0100. From that time it began increasing again, passing $N_{max} = 3000$ el/cm^3 at 0215 and then, probably, continued to increase somewhat more. This Figure is typical of many of the examples we have noted of sporadic *E*.

The other main group of sporadic *E* records is represented by Fig. 3C. In this record it is possible to see reflections from both the 3000-el/cm^3 level and from the sporadic *E* for a fair amount of the time that the sporadic *E* was visible. It should be noted that, on this type of record, the group height of the sporadic *E* is greater than, for example, on Fig. 3A. This would indicate that N_{max} is not below but probably very close to 3000 el/cm^3, causing an appreciable amount of retardation. It would also indicate that the *E*-layer is fairly thin, part of the energy leaking through and part reflected. Somewhat later on, Fig. 3C at about 0225, we see a typical example of the waves completely penetrating the *E*-layer. The corresponding decrease of extra retardation in the *E*-layer is seen to amount to about 18 kilometres.

The above explanation should mean that we are only able to detect sporadic *E* on 150 Kc/s under the condition that the maximum electron density of the *E*-layer drops close to, or below, 3000 el/cm^3. This, on the other hand, means that it is not necessary that the presence of a reflecting stratum slightly above the

E-layer be limited only to disturbed conditions. It could well be there all the time, but normally the E-layer would prevent us from detecting its existence. Measurements on other frequencies at latitudes appreciably South of the Auroral Zone have, in general, shown that sporadic E is not confined to disturbed conditions. On the other hand, as State College is situated but little South of the Auroral Zone, geomagnetic latitude 52°N, it could well be possible that this type of sporadic E is related to some of the forms so frequently noted in, say, Southern Canada. Distant auroral displays have, for example, been visible here on several occasions when sporadic E has been recorded.

However, it still remains to explain why the normal electron density in the E-layer should show such a marked decrease during disturbed conditions. The generally accepted value of the effective recombination coefficient α is, in the E-layer, $\alpha_{day} = 1{\cdot}1 \times 10^{-8}$ (RYDBECK, 1946) and $\alpha_{night} = 2 \times 10^{-9}\text{cm}^3\text{sec}^{-1}$ (APPLETON, 1937). However, as BATES and MASSEY have pointed out (1946), the nocturnal value is probably too low because of the neglect of sources of ionization. They suggest that the true night value is the same as the day value.

We have made an approximate determination of α_{night} during disturbed conditions. Records showing the typical cusp indicating that $N = 3000$ el/cm^3 was reached and, immediately thereafter, showing sporadic E were selected for the investigation. Utilizing the point where $N = 3000$ el/cm^3 and published E-layer data*, it was possible to determine the critical frequency of the E-layer at sunset. The effective recombination coefficient, assumed to be independent of height, could then be computed by solving the well-known equation

$$\frac{dN}{dt} = -\alpha N^2 \tag{2}$$

where

$$\alpha = \alpha_e + \lambda\alpha_i. \tag{3}$$

α_e = true recombination coefficient between electrons and positive ions.

α_i = true recombination coefficient between negative and positive ions.

λ = ratio between negative ions and electrons.

N = the electron density

It was possible to find 25 records where there was no doubt that the maximum electron density decreased below 3000 el/cm^3. The median value of α_{night} obtained from these determinations is $\alpha = 3{\cdot}9 \times 10^{-8}$ cm^{3}sec^{-1}. The minimum value is $1{\cdot}6 \times 10^{-8}$ and the maximum value $7{\cdot}4 \times 10^{-8}$cm^{3}sec^{-1}. The order of magnitude of the effective recombination coefficient, therefore, seems to be at least a factor of ten greater during disturbed conditions than under undisturbed conditions.

It is well known that during a magnetic disturbance a redistribution usually takes place in the F_2-layer; maximum electron density decreases and the thickness increases. A similar behaviour of the E-layer could explain this apparent increase in the recombination coefficient. However, an inspection of the individual records shows this effect to be unlikely. We would expect an appreciable increase in the

* For example, in "Ionospheric Data, Series F," issued by C.R.P.L., Nat. Bur. of Stand., Washington D.C.

group retardation if a redistribution occurred. As this does not happen, we conclude that the increase in the effective recombination coefficient probably is real. The frequent records showing reflections from both the normal E and the sporadic E with a separation in virtual height of only some 15 kilometres also indicate that the shape of the layer is probably unaltered.

As long as we do not have a more complete picture on what is going on during a magnetic disturbance, it is difficult to draw a conclusion as to what actually causes this apparent change in the recombination coefficient. A possible explanation would be to assume, as already suggested by APPLETON and NAISMITH (1940), an impact of high energy corpuscles, able to cause the ionization observed as sporadic E. Due to a similar mechanism, we could expect reactions in the normal E-layer, possibly increasing the value of λ in Eq.(3). However, it is apparent that before anything can be concluded in this direction, we should have some knowledge of, for example, the night sky spectrum during these conditions.

Finally, it should be mentioned that it has not yet been possible to detect any sign of reflections from the F_2-layer on 150 Kc/s. In the above-mentioned cases, when the maximum electron density of the E-layer decreased below 3000 el/cm^3, we would expect F_2-reflections on the records. However, this has never been the case. Every time they seem to have been obscured by the sporadic E-layer developed above the normal E-layer.

VOL.
4
1953

THE EFFECT OF SOLAR FLARES

It is a well known fact that some solar flares give rise to a phenomenon called SID, sudden ionospheric disturbances. The SID's manifest themselves by causing a complete disappearance of high frequency signals received via the ionosphere and also a complete disappearance of echoes received on an ionospheric high frequency recorder. The only possible explanation from this point of view would be the assumption that the radiation from some of the flares would be of such wave-length and such an intensity as to be able to build up a highly absorbing stratum of ionization below the E-layer. Prevailing suggestions place this ionization within the D-layer.

Recently published data from the British workers (BRACEWELL and STRAKER, 1949), shows a very good correlation between practically all solar flares and recordings of 16 Kc/s phase-heights. The main effect is a change in phase height of the order of 30 to over 500 degrees, corresponding to a change in effective height of reflection from 0·9 to 15 kilometres.

These results, namely, that both 16 Kc/s waves and waves of a frequency higher than 1 Mc/s, are affected by solar flares immediately suggest that there should be some effect noted at 150 Kc/s. And, as the British results definitely indicate that the ionization appearing in connection with the flare is located below the E-layer, it is most likely that the effect would be one of increased absorption.

As an approach to determine if such an effect exists, the $h't$-records were checked against all reported high frequency SID's during the period February 1950—March 1952. However, partly due to lack of records and partly due to the considerable daytime absorption of the signal during the summer months, when the

SID's appear most frequently, this survey cannot give more than an indication of the conditions. The numerical results are listed in Table 3.

Table 3. *Result of comparison of SID's and the* 150 *Kc/s virtual height records during February* 1950—*March* 1952

Total No. of reported SID's	*Total No. of SID's when* 150 *Kc/s records are avail.*	*Number of cases showing total absorp. during SID*	*Number of cases showing partial absorp. during SID*	*Number of cases when the overall absorp. was too high*	*Number of cases where no effect was detectable*
198	82 = 100%	19 = 23·2%	23 = 28·0%	37 = 45·1%	3 = 3·7%

From this Table it can be seen that the existing records only cover about 40% of the reported number of fade-outs. Of these 40%, or 82 cases, only 54·9%, or 45 cases, can be used for the investigation. In the 19 cases showing total absorption the effect was very positive. These records showed continuous reflections before and after the SID. Within ± 2 minutes of the time reported for the start of the SID, the 150 Kc/s reflections disappeared suddenly and completely. No change at all in group height could be detected. The time of the return of the echoes corresponded, in most cases, to within a few minutes of the reported end of the SID. On the other hand, the 23 cases reported as showing partial absorption during the SID, started with the same precision as the last group, but had considerable fluctuations in the times of return of the echoes. In some cases the echoes returned within ten minutes after the SID, as observed at short wavelengths, was over but in several cases the no-echo condition could last as long as several hours. As long as the time of onset shows such a remarkable coincidence with the start of an SID, it seems most reasonable to assume that this long period of complete absorption was due to an SID. We do not need much increased electron density below the *E*-layer to account for complete absorption of the *E*-reflections at 150 Kc/s during daytime.

The remaining three cases, when no effect at all was detectable, are difficult to explain. A determination of the reflection heights shows a height of about 85 kilometres in two cases and 94 in the third. The time of day was near noon in all three cases, a time when the 150 Kc/s reflections would be most sensitive to an increase in absorption. It seems, therefore, most likely, in these three cases, that the increased ionization causing the SID occurred at a height above the 3000 electrons/cm^3 level in the *E*-layer.

A direct comparison between reported solar flares and the virtual height records has also been made. The results obtained during the period September 1951—February 1952 are presented in Table IV.

It is interesting to note that the percentage of cases showing a total absorption of the 150 Kc/s reflections was 23·2% both by comparison with solar flares and SID's. This remarkable agreement must be coincidental, but as far as the order of magnitude is concerned, it could be significant; especially as a check of the individual cases showed that almost all of the reported solar flares causing total

Table 4. *Result of a comparison of solar flares and the* 150 *Kc/s virtual height records during September* 1951—*February* 1952

Total No. of reported flares	*Total No. of flares when* 150 *Kc/s records are avail.*	*Number of cases showing total absorp. during flare*	*Number of cases showing partial absorp. during flare*	*Number of cases when the overall absorp. was too high*	*Number of cases where no effect was detectable*
121	69 = 100%	16 = 23·2%	8 = 11·6%	35 = 50·7%	10 = 14·5%

absorption on 150 Kc/s *did not* cause fade-outs at higher frequencies. The fact that solar flares can affect low frequency reflections but not high frequency reflections is also in accordance with the results of the British investigations on 16 Kc/s (BRACEWELL and STRAKER, 1949). However, the number obtained for cases when no effect was noticeable at 150 Kc/s, 14·5%, shows that the 150 Kc/s reflections are a much less sensitive indicator of solar flares than 16 Kc/s reflections.

Sample records showing total absorption during solar flares are depicted in Figs. 1B and 1C. The fade-out depicted in Fig. 1B was recorded on 15 December 1951. The corresponding solar flare is reported during the period 1049–1154 LMST and was of importance 1. No corresponding high frequency fade-out was reported. It can be seen from the record that the time of beginning of the fade-out on 150 Kc/s coincides, within one minute, with the onset of the flare. The recovery of the reflections is somewhat delayed; however, a weak reflection is detectable immediately after the end of the flare.

VOL. 4 1953

In Fig. 1C another fade-out is depicted, this time recorded on 17 September 1951. Two different flares were reported near the time of the fade-out. One flare, of importance 1, was reported as active during 1539–1639 LMST with its maximum at 1552 LMST. The other flare was reported at 1549 LMST and of importance 2. A corresponding high frequency fade-out is reported at 1547–1604 LMST. A comparison with the 150 Kc/s record shows that the low frequency fade-out started at exactly the same time and lasted one minute longer than the high frequency fade-out. Fig. 1C also indicates that the solar flare of intensity 2 is more likely to produce a fade-out than the one of intensity 1. These two flares were located very close together on the sun's disk. The recovery of the fade-out in Fig. 1C is very typical of most of the fade-outs we have observed. The shape of the recovery curve suggests that the reflection is exposed to an extra group retardation which gradually decreases. However, as explained earlier in the paper, a very high degree of absorption would also produce this effect. As this is very likely to be the case, it is safer to discard this piece of information. It is believed that a careful investigation of phase height records obtained on 150 Kc/s could settle this question.

Finally, it must be stressed that, owing to the lack of sufficient midsummer daytime data during the interval of interest, the numbers quoted for the correlations with SID's and solar flares should be regarded as preliminary.

Acknowledgments—The author wishes to express his sincere thanks and appreciation to R. J. NERTNEY, N. DAVIDS, R. E. JONES, A. H. WAYNICK and other

members of the Ionosphere Research Laboratory for their many helpful suggestions and valuable criticism. The author also wishes to thank A. J. HALEY especially for his unfailing interest in supervising the recording equipment.

REFERENCES

APPLETON, E. V.	1937	Regularities and irregularities in the Ionosphere. Proc. Roy Soc. A. **162**, 451–479, October
APPLETON, E. V. and NAISMITH, R.	1940	Normal and abnormal region-*E* ionization. Proc. Phys. Soc. **52**, 402–415
BATES, D. R. and MASSEY, H. S. W.	1946	The basic reactions in the upper atmosphere. I. Proc. Roy. Soc. **187A**, 261–296
BLAIR, J. C., BROWN, J. N. and WATTS, J. M.	1952	A sweep frequency ionospheric recorder for the low frequencies. Paper presented at the U.R.S.I. Meeting in Washington, D.C., April
BRACEWELL, R. N. and STRAKER, T. W.	1949	The study of solar flares by means of very long radio waves. M.N.R.A.S. **109**, No. 1
BROWN, J. N. and WATTS, J. M.	1950	Ionospheric observations at 50 Kc, J. Geophys. Res. **55**, No. 2, June
DAVIDS, N.	1952	Dispersion effects at 150Kc/s for a particular *E*-layer model. J. Atmosph. Terr. Phys. **2**, 324
GARTLEIN, C. W. and MOORE, R. K.	1951	Southern extent of aurora borealis in North America. J. Geophys. Res. **56**, No. 1, March
GERBER, W. and WERTHMULLER, A.	1947	Nahschwundtatigkeit der Rundspruchsender. Technische Mitteilungen der Schveizerischen P.T.T. **XXV**, No. 1
GIBBONS, J. J. and NERTNEY, R. J.	1952a	A method for obtaining the wave solutions of ionospherically reflected long waves, including all variables and their height variation. J. Geophys. Res. **57**, No. 3, September
GIBBONS, J. J. and NERTNEY, R. J.	1952b	Wave solutions, including coupling of ionospherically reflected long radio waves for a particular *E*-region
LINDQUIST, R.	1952	An investigation of the ionizing effect in the *E*-layer near sunrise. J. Geophys. Res. **57**, No. 4, December model. J. Geophys. Res., **57**, No. 3, September
HELLIWELL, R. A. MALLINCHRODT, A. J. KRUSE, F. W., Jr. and WANBSGANSS, B. A.	1950	Pulse studies of the ionosphere at low frequencies. Electr. Res. Lab., Stanford Univ., Calif. (March 15)
HELLIWELL, R. A.	1952	Sporadic-*E* stratification and correlation with low-frequency soundings. Paper presented at the U.R.S.I. Meeting in Washington, D.C. (April)
RYDBECK, O. E. H.	1946	Chalmers Solar Eclipse Ionospheric Expedition 1945. Trans. Chalmers Univ. of Tech., Gothenburg, No. 53
WATTS, J. M. and BROWN, J. N.	1951	Effects of ionosphere disturbances on low-frequency propagation. J. Geophys. Res. **56**, No. 3, September
WHALE, H. A.	1951	Fine structure of the ionospheric region *E*. J. Atmosph. Terr. Phys. **1**, 233

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 28 to 43. Pergamon Press Ltd., London

Winds in the ionospheric regions

S. Deb

Institute of Radio-Physics and Electronics, Calcutta University

(*Received May* 21 1953)

ABSTRACT

The problem of wind circulation in the upper atmosphere, above the 100 km level is investigated by following a method of synthesis by combining theoretical reasonings with experimental data. Starting with the available informations on height distribution of temperature at three different latitudes viz. 30° N, 45° N and 80–90° N, horizontal pressure gradients and, hence, wind velocities at different heights are computed. The effect of kinematic viscosity on the wind velocities is taken into account in general terms. The final results suggest three possible models of wind circulation in the region 100–400 km. In order to judge the relative merits of these models, the results of wind measurement in this region, as are available, are collected and tabulated. It is found that one, at least, of the theoretical models agrees well with the observed wind data. According to this model wind is south-west in the E and F_1 regions of the ionosphere and north-east or east in the F_2 region. The wind velocity rises with height from 50 m.p.h. at 100 km to over 200 m.p.h. above 200 km. There is a seasonal variation in the magnitude of the velocity. This is found to be distinctly lower in summer. The model, however, does not predict any change in direction—diurnal or seasonal—as has been reported by observers. Possible explanations of these discrepancies are discussed. The height distributions of temperature on which the theoretical models are based, are critically examined. Necessity for further rocket observations in different latitudes is emphasized.

VOL.
4
195

1. Introduction

The high regions of the atmosphere—mesophere and the thermosphere—are known to be subject to winds and turbulence like the low tropospheric regions (Mitra, 1952). The various upper atmospheric phenomena and observations that have led to this conclusion are briefly as follows: Movements of noctilucent clouds and of meteor trails and observations on the so-called sporadic E region of the ionosphere indicate the existence winds of considerable velocity in the region 80–120 km; variations in the intensity of radio noises of extra-terrestrial origin and of the night air-glow show that regions much higher—regions as high as the F_2 region of the ionosphere (that is, up to a height of 250–300 km)—are also subject to winds. The study of these wind phenomena in the high atmosphere is not only of geophysical interest but is also of considerable strategic importance.

Some of the investigations—experimental and theoretical—that have been made in this subject in recent years may be mentioned here: S. N. Mitra (1949), Sheppard (1949) and others have collected data for the heights 50 km to 100 km and have attempted to deduce therefrom the trend of wind velocities in different hours of the day and in the different seasons of the year in this region. Chiu (1951) has attempted to infer the wind system theoretically from a consideration of the physical state of the upper atmosphere. He has discussed in detail the factors which are likely to control the pattern of air movement in the upper regions and has tried to find out, in a qualitative way, the salient features of the circulatory system. Yerg (1951) has shown that there may be a wind circulation above and below the 200 km level, where according to him the viscosity attains a maximum value.

A fruitful synthetic method of investigating the probable wind circulation in

the upper atmospheric regions is one involving a combination of experimental results and theoretical reasonings. Such a method has been adopted by KELLOGG and SCHILLING (1951). These authors, starting with the available data of wind velocities and pressure distribution up to the height of 100 km, have computed the possible height distribution of temperature at various latitudes and have tried to deduce therefrom a model circulation system as would agree with the largest number of experimental observations. No analogous study appears, however, to have been made of the region above 100 km, which, as already mentioned, is also subject to winds. This study is proposed to be made in the present paper, following a method of synthesis similar to that of KELLOGG and SCHILLING; there will, however, be the difference that the starting point will be the data, available as they are, of the height distributions of pressure and temperature in the different latitudes. These data will be utilized for computing theoretically the model wind circulation in the region 100 km to 300 km (*E* and *F* regions of the ionosphere), and, discrepancies as may be found between this computed model and the experimentally observed wind velocities will then be discussed.

2. HEIGHT AND LATITUDINAL DISTRIBUTION OF TEMPERATURE

The most comprehensive review of the available information regarding temperature distribution in the high atmosphere is that made recently by GERSON (1951). Some plausible estimates of temperature have also been made from the rocket exploration of the upper atmosphere (1952). In this Section attempt will be made to construct, as best as possible from the available data, the temperature profile in the height range 100 to 400 km at the three latitudes 90°–80° N, 45° N and 30° N.

(i) *Latitude* 90°–80° N. A complete height distribution of temperature for the region above 100 km for winter, for the north polar region has been given by GERSON. This has been deduced from the values of scale heights as obtained by HARANG from auroral observations. This distribution is reproduced in Fig. 1, curve I. No data are, however, given by HARANG for summer. We can nevertheless make a rough estimate for the distribution for this season from the following consideration:

At 80° latitude, the region above 270 km is illuminated by solar rays both in winter and in summer. Hence, we may assume that the temperature distribution in this region does not change much from summer to winter. The winter temperature as obtained from HARANG's data may, therefore, be taken as representing approximately, the summer conditions as well for the region above 270 km.

For the region below 270 km, we note that according to KELLOGG and SCHILLING (1951) the temperature in summer at the 100 km level is of the order 250°K. We may assume—as is generally done—that a linear rise of temperature exists above this level. If the gradient is assumed to be such as to lead to HARANG's value (2250°K) at 270 km level, the temperature distribution resulting from this assumption, combined with HARANG's results above 270 km level, gives a picture of the probable temperature distribution in summer above 100 km level. This is shown in Fig. 1, curve II.

(ii) *Latitude* 45° N. From a careful study of the data given by various workers, GERSON has adopted model height distributions of temperature for both summer

and winter conditions, for latitude 45° N. These are reproduced in Fig. 1, curves III and IV.

It is to be noted that according to GERSON the values given represent the upper limits of the values as given by other workers.

(iii) *Latitude* 30° N. Height distribution of temperature for this latitude has recently been estimated by the Rocket Panel of the Harvard University (1952). The estimate, however, gives only the average annual temperature for an atmosphere whose constituents (N_2 and O_2) are undissociated. Further, the estimate is

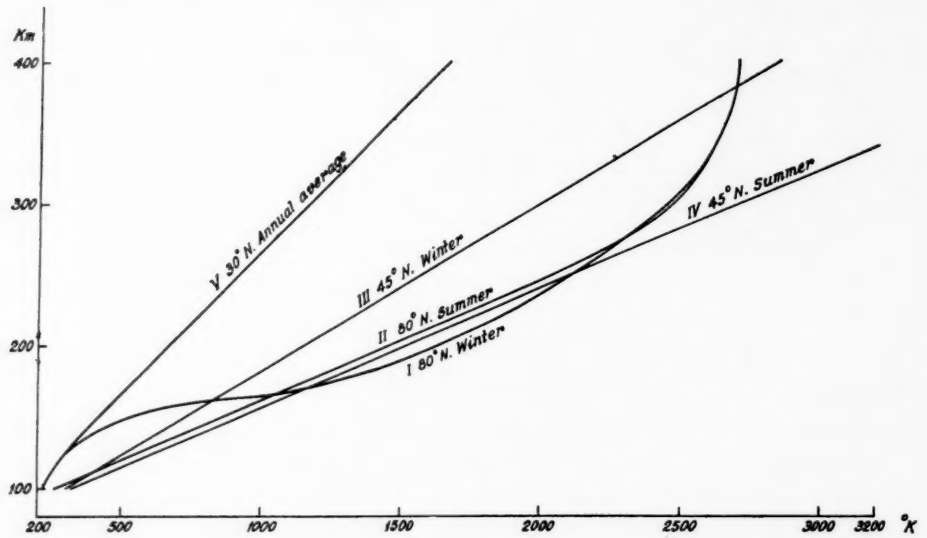


Fig. 1. Showing the height distribution of temperature at latitudes 90–80° N, 45° N and 30° N. Adapted from GERSON's review (1951) and the data given by the Rocket Panel, Harvard University (1952).

only up to a height of 220 km. It is stated that considerable uncertainties exist for the data for the higher levels. Rocket data have also been utilized by HAVENS, KOLL and LAGOW (1952) to compute the temperature for the region 100–160 km. These authors, however, assume that the atmospheric constituents in this region are O and N_2. This assumption we know is in accord with current ideas regarding dissociation of O_2 in the upper atmosphere. The temperature distribution as obtained by HAVENS *et al*, may, therefore, be taken as reasonably accurate, at least up to the 200 km level (DEB, 1952; MOSSES and TA-YOU-WU, 1952). It is, therefore, possible to recompute, with the help of the data given by the Rocket Panel, the temperature distribution up to the 220 km level. The temperature values obtained show an approximately linear rise with height. For purposes of our investigation on wind circulation we will assume this rising temperature gradient to persist for the region above 220 km as well. The height distribution obtained on these assumptions is shown in Fig. 1, curve V.

It is to be noted, in this connection, that nitrogen has been shown to dissociate in significant proportion in the F region (DEB, 1952). The effect of this would be to

VOL. 4 1953

lower somewhat the temperature in this region. However, in view of the large uncertainty in our estimate of temperature in the lower regions—from which the temperature data in the F region have been deduced—no point would possibly be gained at this stage by taking this effect into account.

3. Method of Calculation and Results

The height distribution of temperature, in the high atmospheric regions under consideration, being different in different latitudes (as represented by Fig. 1), the corresponding pressure distributions will also be different. The isobaric surfaces will, therefore, in general, be inclined to the horizontal and, hence, wind circulations will tend to develop—the direction of the winds being along the pressure gradient. Due to the earth's rotation, however, the wind will also be subject to the Coriolis force. The effect of this will be to deflect the wind in a direction parallel to the isobaric surface. For a steady circulation to exist the forces due to the pressure gradient and to the earth's rotation must balance each other. Under such condition a south to north flow is turned into a west to east flow and a north to south flow into an east to west flow.

The wind velocity V_ω in the case of steady circulation, when viscous forces are absent, is given for latitude ϕ by

$$V_\omega = \frac{g \tan \theta}{2 \omega \sin \phi}, \tag{1}$$

where θ is the angular inclination of the isobaric surface to the horizontal, ω is the angular velocity of the earth's axial rotation ($= 2\pi/86164$ radians/sec) and g is the accleration due to gravity. If frictional forces are operative, a component of the wind velocity along the normal to the isobaric surfaces will appear. Further, such forces will preclude the possibility of the existence of any large scale velocity gradient from layer to layer.

To determine the wind system we have, therefore, to first construct the isobaric surfaces. This can be done by calculating the pressure distribution with the help of the temperature diagram of Fig. 1 and the hydrostatic equation

$$\frac{\mathrm{d}P}{P} = -\frac{mg}{kT}\,\mathrm{d}h = -\frac{1}{H}\,\mathrm{d}h.$$

$$\text{Integrating, } P = P_0 \exp\left[-\int_0^h \frac{\mathrm{d}h}{H}\right], \tag{2}$$

where P, T and m are the pressure, temperature and mean molecular weight at height h and P_0 is the pressure at some datum level. H is the so-called scale height of the atmosphere at height h and k is the Boltzmann constant.

The various parameters required in the calculation, namely, (i) pressure at the datum level, (ii) mean molecular weight of air, (iii) values of g and (iv) value of viscosity, are now obtained as follows:

(i) *Pressure at the datum level* (100 *km*). The datum level is taken as the 100 km level. The pressure at this level for the polar region may be obtained from the

data given by KELLOGG and SCHILLING (1951). For 45° N latitude the pressure has been estimated by GERSON (1951) by assuming that the atmosphere there is composed mainly of N_2 and O. For 30° N latitude we assume the value given by HAVENS, KOLL and LAGOW (1952). These data are collected in Table 1.

Table 1. *Values of pressure at the datum level* (100 *km*)

30° N *Mean annual*	45° N		80°–90° N	
	Summer	*Winter*	*Summer*	*Winter*
$5{\cdot}36 \times 10^{-4}$	$1{\cdot}488 \times 10^{-3}$	$1{\cdot}488 \times 10^{-3}$	$8{\cdot}5 \times 10^{-4}$	$4{\cdot}5 \times 10^{-4}$

The values are given in mb (1 mb = 0·750 mm)

(ii) *Mean molecular weight of air.* It is now well recognized that oxygen is almost completely dissociated above the 100 km level (MOSSES *et al.* 1952). It has further been shown that nitrogen also is dissociated in significant proportion above the 200 km level. The latter dissociation, however, increases very slowly with height, so that even at a height of 400 km it is at most 25–30 per cent. On the basis of the above, we obtain the following average values of the mean molecular weight (m) of air:

$$m = 23{\cdot}95 \text{ in the range } 100\text{–}200 \text{ km.}$$

$$m = 22{\cdot}28 \text{ in the range } 200\text{–}400 \text{ km.}$$

(iii) *Acceleration due to gravity.* This is easily computed—neglecting the variation with latitude—by using the inverse square law.

(iv) *Viscosity.* According to a recent calculation kinematic viscosity plays an effective rôle in the region 90–300 km, its value varying between 10^6 and 10^3 (YERG, 1952). This is precisely the region in which we are interested. Hence, the wind velocity in this region may always be assumed to have a component normal to the isobaric surfaces and to be subject to appreciable retarding action.

With the numerical values deduced as above and with the height distribution of temperature as obtained from Fig. 1, we can calculate with the help of Eq(2) the values of pressure at suitable height intervals. The values so determined, at intervals of 20 km, are shown in Table 2.

It is seen from the Table that the computed pressures do not give any coherent picture of the distribution of the isobaric surfaces. In particular, the data obtained for latitude 30° N from V_2 rocket observations are abnormally low compared to those for higher latitudes. This lack of compatability is obviously due to the inadequacy and inaccuracy of the observational data. Nevertheless, we can make use of the data in Table 2 for examining the pattern of air movement as would result from isobaric surfaces computed from the height distributions of pressure taken in pairs, viz. pressure data for 90°–80° and 45°, for 90°–80° and 30°, and for 45° and 30° N. The pattern of air movements thus deduced may then be compared with observed wind data (Sec. 4) to check which pattern or which temperature distribution represents most closely the actual state of affairs. It is to be

VOL. 4 1953

OL.
953

Table 2. Computed height distribution of pressures for latitudes 30° N, 45° N and 80–90° N

Height km	*Pressures in mb*						
	30°N	45° N			80–90° N		
	Mean annual	*Summer*	*Winter*	*Mean annual*	*Summer*	*Winter*	*Mean annual*
100	$5{\cdot}36 \times 10^{-4}$	$1{\cdot}488 \times 10^{-3}$	$1{\cdot}488 \times 10^{-3}$	$1{\cdot}488 \times 10^{-3}$	$8{\cdot}5 \times 10^{-4}$	$4{\cdot}5 \times 10^{-4}$	$6{\cdot}5 \times 10^{-4}$
120	$4{\cdot}653 \times 10^{-5}$	$4{\cdot}14 \times 10^{-4}$	$3{\cdot}63 \times 10^{-4}$	$3{\cdot}88 \times 10^{-4}$	$1{\cdot}92 \times 10^{-4}$	5×10^{-5}	$1{\cdot}21 \times 10^{-4}$
140	$9{\cdot}32 \times 10^{-6}$	$1{\cdot}85 \times 10^{-4}$	$1{\cdot}39 \times 10^{-4}$	$1{\cdot}62 \times 10^{-4}$	$7{\cdot}84 \times 10^{-5}$	10^{-5}	$3{\cdot}92 \times 10^{-5}$
160	$3{\cdot}3 \times 10^{-6}$	$1{\cdot}03 \times 10^{-4}$	$6{\cdot}74 \times 10^{-5}$	$8{\cdot}52 \times 10^{-5}$	$4{\cdot}15 \times 10^{-5}$	4×10^{-6}	$2{\cdot}28 \times 10^{-5}$
180	$1{\cdot}176 \times 10^{-6}$	$6{\cdot}44 \times 10^{-5}$	$3{\cdot}76 \times 10^{-5}$	$5{\cdot}1 \times 10^{-5}$	$2{\cdot}53 \times 10^{-5}$	$2{\cdot}35 \times 10^{-6}$	$1{\cdot}38 \times 10^{-5}$
200	$5{\cdot}15 \times 10^{-7}$	$4{\cdot}39 \times 10^{-5}$	$2{\cdot}3 \times 10^{-5}$	$3{\cdot}35 \times 10^{-5}$	$1{\cdot}69 \times 10^{-5}$	$1{\cdot}62 \times 10^{-6}$	$9{\cdot}26 \times 10^{-6}$
220	$2{\cdot}59 \times 10^{-7}$	$3{\cdot}23 \times 10^{-5}$	$1{\cdot}56 \times 10^{-5}$	$2{\cdot}4 \times 10^{-5}$	$1{\cdot}23 \times 10^{-5}$	$1{\cdot}07 \times 10^{-6}$	$6{\cdot}69 \times 10^{-6}$
240	$1{\cdot}41 \times 10^{-7}$	$2{\cdot}5 \times 10^{-5}$	$1{\cdot}1 \times 10^{-5}$	$1{\cdot}8 \times 10^{-5}$	$9{\cdot}35 \times 10^{-6}$	$8{\cdot}3 \times 10^{-7}$	$5{\cdot}09 \times 10^{-6}$
260	$8{\cdot}28 \times 10^{-8}$	$1{\cdot}98 \times 10^{-5}$	$8{\cdot}13 \times 10^{-6}$	$1{\cdot}4 \times 10^{-5}$	$7{\cdot}51 \times 10^{-6}$	$6{\cdot}568 \times 10^{-7}$	$4{\cdot}09 \times 10^{-6}$
280	$5{\cdot}15 \times 10^{-8}$	$1{\cdot}61 \times 10^{-5}$	$6{\cdot}19 \times 10^{-6}$	$1{\cdot}11 \times 10^{-5}$	$5{\cdot}93 \times 10^{-6}$	$5{\cdot}29 \times 10^{-7}$	$3{\cdot}23 \times 10^{-6}$
300	$3{\cdot}25 \times 10^{-8}$	$1{\cdot}33 \times 10^{-5}$	$4{\cdot}82 \times 10^{-6}$	$9{\cdot}06 \times 10^{-6}$	$4{\cdot}84 \times 10^{-6}$	$4{\cdot}32 \times 10^{-7}$	$2{\cdot}64 \times 10^{-6}$
320	$2{\cdot}24 \times 10^{-8}$	$1{\cdot}12 \times 10^{-5}$	$3{\cdot}84 \times 10^{-6}$	$7{\cdot}52 \times 10^{-6}$	$3{\cdot}99 \times 10^{-6}$	$3{\cdot}56 \times 10^{-7}$	$2{\cdot}28 \times 10^{-6}$
340	$1{\cdot}56 \times 10^{-8}$	$9{\cdot}61 \times 10^{-5}$	$3{\cdot}11 \times 10^{-6}$	$6{\cdot}36 \times 10^{-6}$	$3{\cdot}33 \times 10^{-6}$	$2{\cdot}956 \times 10^{-7}$	$1{\cdot}82 \times 10^{-6}$
360	$1{\cdot}116 \times 10^{-8}$	$8{\cdot}32 \times 10^{-6}$	$2{\cdot}56 \times 10^{-6}$	$5{\cdot}44 \times 10^{-6}$	$2{\cdot}77 \times 10^{-6}$	$2{\cdot}465 \times 10^{-7}$	$1{\cdot}51 \times 10^{-6}$
380	$8{\cdot}06 \times 10^{-8}$	$7{\cdot}28 \times 10^{-6}$	$2{\cdot}13 \times 10^{-8}$	$4{\cdot}7 \times 10^{-6}$	$2{\cdot}23 \times 10^{-6}$	$2{\cdot}064 \times 10^{-7}$	$1{\cdot}27 \times 10^{-6}$
400	6×10^{-9}	$6{\cdot}42 \times 10^{-6}$	$1{\cdot}8 \times 10^{-6}$	$4{\cdot}11 \times 10^{-6}$	$1{\cdot}95 \times 10^{-6}$	$1{\cdot}733 \times 10^{-7}$	$1{\cdot}06 \times 10^{-6}$

noted that the results to be described relate exclusively to the middle latitudes of the northern hemisphere. Results for the southern hemisphere may be a mirror image or a continuation of the pattern in the northern hemisphere. However, sufficient data are not yet available for even an approximate deduction on this point.

(i) *Latitude* 90°–80° *and* 45° *N*. An examination of the computed height distributions of pressure at these latitudes show that the pressure tends to increase towards the equator. As such the isobaric lines tends to slope towards the pole. At 105 km for example, this slope is found to be 1/508·7 and $g = 947$. One thus obtains from Eq (1) $V_\omega = 200$ m.p.h. for latitude 60° N. Detailed calculations reveal the following:

A south to north movement of air exists all the year round in the region 100 km and above. No evidence of any seasonal change in direction exists for this region.

Wind velocity is distinctly lower in summer. Thus, in the region round 100 km, the speed in summer is 50–60 m.p.h. and is approximately 100 m.p.h. in winter.

Strong vertical gradient of wind velocity tends to develop above the 100 km level. Thus, in winter the velocity increases from about 100 m.p.h. at the 100 km level to more than 400 m.p.h. at 120 km level, and to several thousand m.p.h. at 100 km level. However, viscous forces also rise to a maximum value in the 120–200 km region leading to a rapid decrease of the gradient with height.

As a result of the poleward transfer of air and the damping effect due to viscosity, return currents of larger velocity may be established at about 250 km level in summer and at a somewhat higher level in winter. Air movement in the F_2 region of the ionosphere is, therefore, likely to be from south to north.

VOL. 4 1953

Note: This return current may be either through a lower region or through a region higher up in the *F* layer. The velocity of the return flow in the former case need be small whereas in the latter case it should be large. From the computed pressure distribution, no evidence of a return current through the higher region is found to exist. It should, however, be noted that the computation for higher region is based on very uncertain data. Results of theoretical and experimental investigations indicate that the composition and physical states of the *E* and the *F* regions of the atmosphere are widely different. A large scale exchange of air extending from 100 to the 400 km level, therefore, seems to be highly improbable. Accumulation of air in the polar region resulting from flow in and immediately above the *E* layer, may be compensated by an equatorward flow below at about 80 km level. The accumulation resulting from flow higher up may, however, likely be balanced by a return flow above the 200 km level.

If the effects due to Coriolis force and viscosity are taken into account, one arrives at the following conclusions: In northern hemisphere a SW wind having a speed of 50–200 m.p.h. exists in the *E* region of the ionosphere and a NE wind having a larger speed exists in the lower strata of the *F* region of the ionosphere. Higher up E wind may prevail due to fall in the value of viscosity.

The circulation pattern described above is shown in Fig. 2 (i).

(ii) *Latitude* 90°–80° *and* 30°. The following characteristics of the wind system are obtained from an examination of the computed height distributions of pressure at these latitude.

There is indication of a feeble seasonal change in wind direction in the region 100–120 km. Air movement is from south to north in winter and from north to

south in summer, the latter being in preponderance. No indication of any such change in wind direction exists for the region above 150 km, for which the direction of air movement is always from north to south.

At the 100 km level the wind velocity is higher in summer. It is about 90 m.p.h. in this season and 40 m.p.h. in winter. At the F region level the difference is more marked.

In winter the vertical gradient of wind velocity tends to be small in the region 100–180 km and appreciable in the region above. In summer, the gradient tends

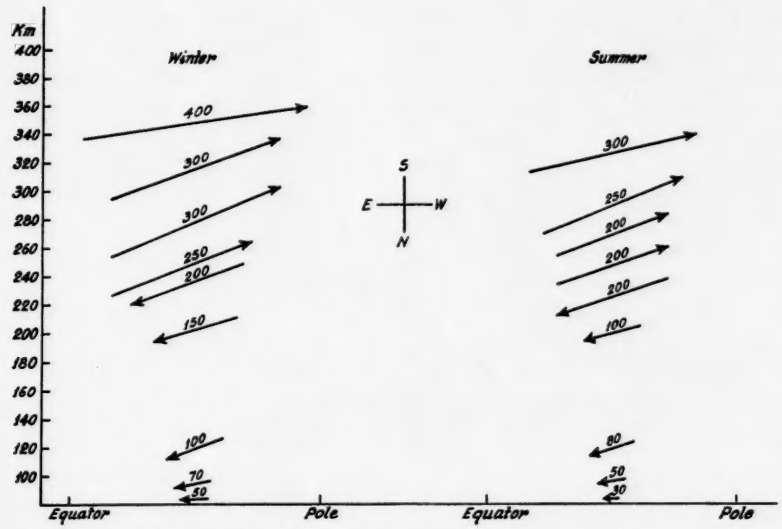


Fig. 2 (i). Showing the possible wind circulation system (above the 100 km level) as derived theoretically from a consideration of the temperature distributions at the latitudes 90–80° N and 45° N.

to be moderate in the former region and extremely large in the latter. This last tendency above 180 km level is, however, checked by the marked effect of viscous forces operative at that region.

As a result of the rapid equatorward transfer of air, and damping effect due to viscosity a return current would probably be established at about 250 km in winter and at some lower height in summer. Air movement in the F_2 region may be, therefore, from south to north all the year round.

If the effects due to Coriolis force and viscosity are taken into account, then the following conclusions may be drawn: The wind in the E region is SW in winter and is very feeble; it is NE in summer with a speed of above 100 m.p.h. In the F_2 region the wind is always SW or W with a speed 200–300 m.p.h.

The circulation pattern described above is shown in Fig. 2 (ii).

(iii) *Latitude* 45° *and* 30°. The following characteristics are obtained from an examination of the computed height distributions of pressure at these latitudes:

The direction of air movement is from north to south all the year round.

Wind velocity is of the same order both in summer and in winter.

Indication of strong vertical gradient of wind speed exists. A value as high as several thousands of miles per hour may be attained even at as low a height as 150 km. Retarding forces due to viscosity will, however, considerably reduce this tendency.

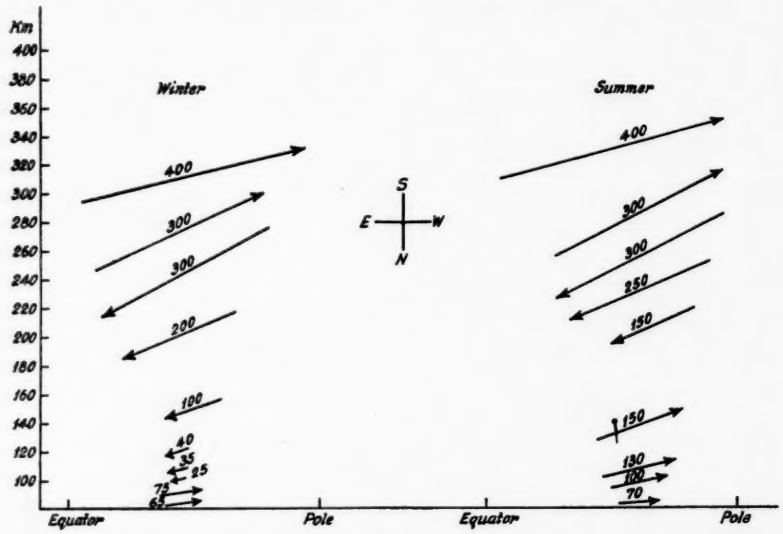


Fig. 2 (ii). Showing the possible wind circulation system (above the 100 km level) as derived theoretically from a consideration of the temperature distributions at latitudes 90–80° N and 30° N.

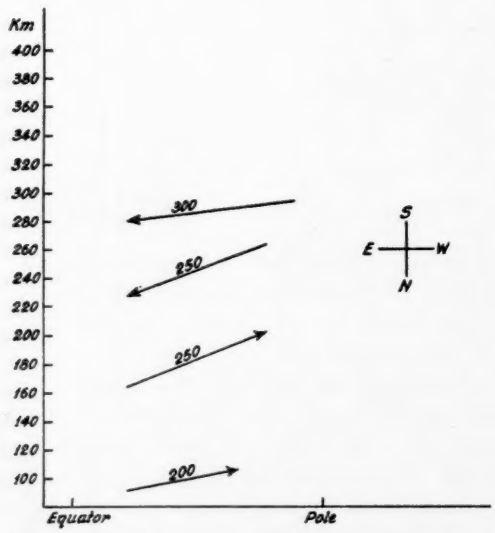


Fig. 2 (iii). Showing the possible wind circulation system (above the 100 km level) as derived theoretically from a consideration of the temperature distributions at latitudes 45° N and 30° N.

As a result of equatorward transfer of air and of the slowing down of speed, a return flow is set up probably above 250 km.

If the effects of Coriolis force and viscosity are taken into account the following conclusions are reached: Wind in the E region is NE with a speed of 200 m.p.h. and in the F_1 region NE with a speed of 250 m.p.h. In the F_2 region and above wind is likely to be due to the return flow which may be SW or, on account of the lowering of viscosity in that region, purely zonal i.e. from W to E.

The pattern of wind circulation described above is shown in Fig. 2 (iii). It is to be noted that the wind patterns given in Fig. 2 for the region 70–90 km are taken from the results obtained by KELLOGG and SCHILLING (1951).

VOL. 4 195

4. Results of Wind Measurement in the Ionosphere

The deductions made in the previous section regarding the probable wind circulation pattern in the regions above 100 km will now be compared with the wind measurement data in the same region. This will enable us to find out which of the calculated patterns is most likely to represent the actual circulation system.

The wind measurement data for the region above 100 km as are available are collected and arranged in Table 3. In columns 1 and 2 the height and the latitude of observation are given. The method of observation and the trends of wind direction and velocity are indicated briefly in columns 3 and 4. Column 4 also contains references to original literature. Wind is designated by its *direction of arrival*. Thus, a north to south wind is designated as "N wind," east to west wind as "E wind" and so on. Velocities are expressed in miles per hour.

It will be noticed from the table that a large proportion of the data is based on ionospheric observations. Since such observations are confined mostly to regions of maximum ionization, the wind data are also available only for such regions, that is, for regions round E and F maxima. Further, the bulk of the available data relates to conditions in the middle and higher latitude of the northern hemisphere. Comparatively few data are available for the southern hemisphere.

The recorded data also show that anomalies exist not only in the magnitude of the observed wind velocity—so that there is an appreciable spread—but what is more, even the directions indicated are sometimes opposite to each other for the same time and place. Thus, while observation No. 4 indicates that in the region 100–120 km, E wind prevails in the morning, observation No. 11 suggests that in the same region W wind prevails at the same hour. However, closer study of the table suggests some definite features of the general trend of the wind circulation in the ionospheric regions. These are as follows:

Region 100–200 *km.*

(i) The wind speed in the region of 100–120 km lies in the range 70–200 m.p.h.

(ii) There exists a tendency of diurnal reversal of wind direction in this region at about midnight. Most of the observations suggest a SW wind from noon to midnight and a NE wind from midnight to noon.

(iii) There also probably exist seasonal changes in wind direction and speed. In summer the speed is lower, but it is more regular. Evidences regarding the seasonal change of direction are not, however, conclusive.

Region 250 *km and above.*

(iv) The wind speed in the region 250 km and above lies in the range 150-250 m.p.h.

(v) There is no clear indication of the existence of a regular diurnal change of the wind direction in this region.

(vi) Positive indications exist of a seasonal change of the wind direction at equinoxes in the F layer. There prevails regular E wind in summer in the northern hemisphere. In southern hemisphere W wind is more frequent in summer.

Table 3. Results of wind measurement above 100 km

Height	*Latitude*	*Method of Observation*	*Wind Speed and Direction*	*Reference*
1. 80–120 km	50° N	Meteor trails and noctilucent clouds	E wind below 80 km. Irregular, with W wind component prominent between 90–100 km. SW wind at 120 km. Speed 224 m.p.h. in summer and 605 m.p.h. in winter	SHEPPARD, 1949
2. 80–150 km	—	Meteor observations	W wind predominant. Speed 129 m.p.h.	WHIPPLE, 1943
3. 90–110 km	25° N	Radio observations of meteors	S.SW and N wind. Speed 70 m.p.h. in summer. Indication of strong wind shear	MANNING, VILLARD, PETERSON, 1950
4. 100–120 km	40° N	Radio observations of sporadic *E*	E wind in morning, W wind in afternoon	FENNEL, 1944
5. 100–120 km	40° N		E wind in winter; speed 292 m.p.h. S.SW wind in summer; speed 90 m.p.h.	FERREL, 1948
6. 100 km	50°–60° N		216 m.p.h.	MEEK, 1949
7. 100–120 km	50°–60° N		Speed 151 m.p.h. Regular diurnal and seasonal changes in wind direction	PHILLIPS, 1952
8. 100–120 km	50°–60° N		Mainly E wind. Speed 78–180 m.p.h.	GERSON, 1951, 1952
9. 100 km *E* layer	52° N		Diurnal change in wind direction. NE wind from midnight to noon. SW wind from noon to midnight. Speed 156 m.p.h.	Geophys. Res. Dir., Record of Colloquium, 1950, MITRA, 1949
10. *E* layer	34·5° N		Semi-diurnal change in wind direction from S to N wind at 07 and 19 hrs. Variation dependent on tidal motion	MATSUSHITA, 1950
11. 100 km	52° N		Regular wind in summer; irregular in winter. E wind at night and W wind at midday. Speed 150 m.p.h.	SALZBERG and GREENSTONE, 1951
12. 110 km	40°–45° N	Radio observations of movements of layer irregularities	N and E wind predominant. Semi-diurnal variation of wind speed with amplitude 80 m.p.h. Positive correlation of wind speed and ionospheric disturbance	CHAPMAN, 1948
13. *F* layer	50° N		SE wind. Speed 112 m.p.h.	Geophys. Res. Dir., Record of Colloquium, 1950; MITRA, 1949
14. F_2 layer	50° N		No diurnal variation. Speed 648 m.p.h.	Geophys. Res. Dir., Record of Colloquium, 1950; KELLOGG and SCHILLING, 1951
15. 250 km	30° S		Mostly W wind. Speed 214–438 m.p.h. Seasonal change at equinoxes. No indication of diurnal change	BEYNON, 1948; MUNRO, 1950; Report on Geophys. Discussion, 1951
16. 250 km	—		W wind. Speed 70–700 m.p.h.	Report on Geophys. Discussion, 1951
17. 250 km	52° N		Speed 160 m.p.h.	Report on Geophys. Discussion, 1951
18. *F* layer	34·5° N		Semi-diurnal change in wind direction from S to N wind at 07 and 19 hrs. Average speed 320 m.p.h.	MATSUSHITA, 1950
19. *F* layer	40°–45° N		Wind speed about 200 m.p.h. No positive indication of any diurnal change	CHAPMAN, 1953
20. 300 km	34·5° N	Observations of night air-glow	N-S.SE wind. Speed 600 m.p.h.	HURUHATA, 1950
21. 300 km	34·5° N		N.NE wind; speed 250 m.p.h. and S.SW wind: speed 200 m.p.h. No regular hour or season of change in wind direction. No correlation with tides	HURUHATA, 1950
22. *E* and *F* layer	30°–40° N	Radio observations of ionospheric ionization	N wind in *E* layers and S wind in F_2 layer. Speed 60 m.p.h.	YUMURA, 1952
23. F_2 layer	60° N	Observations of stellar radio noise from Cygnus and Cassiopeia	Steady E winds (15° N of W) in summer. Speed 150 m.p.h. Tendency of reversal at midnight	MAXWELL and LITTLE, 1952
24. F_2 layer	53° N		Steady E wind (150° S of W) in summer; speed 260 m.p.h. Tendency of reversal at midnight	MAXWELL and LITTLE, 1952
25. 300 km	52° N		E wind. Speed 219 m.p.h. Direction sometimes changed rapidly	RYLE and HEWISH, 1950

VOL. 4 195

5. Discussion

We have now to examine how far the circulation patterns of Fig 2, computed from the pressure distribution data, conform to the patterns described in Sec. 4. In doing so we must, however, remember that the computed patterns can not be expected to agree closely with the observed ones, because of the neglect of many factors which are likely to affect the magnitude and direction of the winds. These are: (i) tidal movements of air, (ii) effect of terrestrial magnetic field. (This is because wind velocities are obtained mostly by radio sounding of the ionosphere, and the terrestrial magnetic field—as it influences the motions of ion and electrons —must also influence the wind observational data), (iii) the changing pressure field caused by the changing height gradient of temperature as may influence the circulation system to some extent, (iv) strong vertical gradient of temperature as may produce turbulence and eddy diffusion. Of these, we have some definite knowledge of the tidal motions. It is known that the tidal motions of air particles are mainly horizontal, and according to the Taylor-Pekeris theory, the velocity attained by the particles at 100 km level is of the same order as those given in Table 3. Wind circulations as developed by the tidal forces are thus superposed on those due to temperature gradients. It is, in fact, well-known that the semi-diurnal barometric pressure oscillations as observed at the ground are due, nearly in equal measures, to tidal forces and to temperature gradients (Mitra, 1952). In examining the discrepancies between the wind patterns as calculated from pressure data and those as deduced from direct observational data, this possible effect of tidal motions has to be kept in mind.

Let us now examine each of the circulation patterns individually.

Between Lat. 90°–80° N and 45° N, Fig. 2 (i). We note that the magnitudes of the computed wind velocities at different levels (Sec. 3) are in very good agreement with the observational results (Sec. 4). The computations also predict a lower but more stable wind flow in summer as actually observed. Further, the computed wind direction in summer is in agreement with observed direction. However, the pattern does not indicate any large-scale diurnal or seasonal reversal of wind direction as required by observational data.

Between Lat 90°–80° N and 30° N, Fig. 2 (ii). The computations predict a larger wind speed in summer. This is contrary to the observed data. The magnitudes of the wind velocities in summer, for both the E and F regions, agree well with observational data, but the direction of the flow is just the reverse. There is again no indication in the computed pattern of the existence of any diurnal or seasonal reversal of wind direction.

Between 45° N and 30° N, Fig. 2 (iii). The predicted wind speeds are a little on the high side, see Sec. 3. Also, there is no large scale difference in the predicted values between summer and winter. Further, the wind direction for the F_2 region is mostly opposite to that observed. As in the two previous cases, there is neither any diurnal change in the wind direction in the E region nor any seasonal change in the F region.

From the comparative study made above it would appear that the wind pattern as shown in Fig. 2 (i) agrees best with the observed wind data. But, before this pattern is accepted as approximating the actual state of affairs, one

has to explain one important discrepancy between observed and computed data: While all the observed data indicate diurnal and seasonal changes in the wind directions, no such corresponding changes are predicted by the computed data. It is to be remembered, in this connection, that to produce such reversal of wind direction, large changes in temperature distribution—so large as to reverse the direction of the gradient—have to be assumed. Such large changes—diurnal or seasonal—are most unlikely. In fact, according to all available informations, the diurnal variations of temperature in the high regions that have been observed are all small (GERSON, 1951).

It has been suggested (Report on Geophys. Discussion, 1951) that so far as the diurnal changes are concerned, they may be an apparent effect only, arising out of the observations being made at different heights at different hours of day and night. This suggestion readily fits in with the wind pattern shown in Fig. 2 (i). It is seen from the figure that within two narrow ranges of height at 80–90 km and at 250 km there are changes in wind direction due to the development of return flow. These are also the heights at which most of the wind observations have been made. For the 80–90 km region a change from NE wind at morning to SW wind at evening might, therefore, have been observed due to a rise in the height of observation (the height of the ionospheric layer) at night by about 10–20 km. Alternatively, it may be supposed that the observed diurnal change in direction is a real effect and is caused by semi-diurnal tidal reversal of flow of the air particles.

Regarding seasonal changes of wind in the higher regions we note from Fig. 2(i) that the region in which the wind changes direction in the F layer lies higher up in winter. This may, at least, partly account for the reported transition from E wind in summer to W wind in winter in the 250 km region. For the polar and equatorial regions an alternative explanation may also be thought of. In the return current above the F_1 region cold air will be moving southward from pole to equator. This might lead to a decrease in temperature with height in that region. YERG (1951) has put forward the view that the consequent lapse rate may be strong enough to break up the whole circulation system into three parts as shown in Fig. 3 (after YERG). It will be seen that in the central region the circulation is composed of W wind below and E wind above and *vice versa* for the polar and equatorial regions. It may, however, be noted that the development of such a complex circulation is possible only when the poleward decrease of temperature is very marked. As such, the circulation may exist only in winter when the region in and around the 200 km level is in complete darkness. In summer months a simple circulation extending from pole to equator is more likely. A transition to this latter pattern from the one shown in Fig. 3, might give rise to a seasonal change in wind direction in the polar and equatorial regions.

The above discussion should not, however, be taken to imply that the height distributions of temperature, as assumed for the latitudes 90°–80° N and 45° N, represent accurately the actual state of affairs. There are several features of the assumed distribution which are open to question. For example, the assumed distribution requires that the temperature within the 170–300 km range should increase poleward in winter. This is difficult to explain in view of the fact that

VOL.
4
1953

in the polar regions the layer up to the 300 km level remains in complete darkness for months together in winter.

Of the various temperature distributions assumed, that for 30° N latitude deserve special consideration. This is because the computations are based on rocket data as have been obtained by direct recordings. The computed data may, therefore, also be regarded as obtained from direct observations. The most important feature of this distribution is the value of temperature in the ionospheric

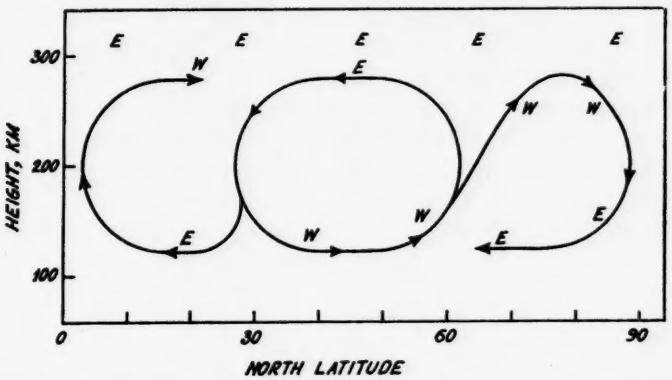


Fig. 3. Possible wind circulation in the F region when the temperature decreases markedly with height. (After YERG.)

region. This is low compared to most of the values estimated previously from indirect observations. However, in the present work the wind velocities have been derived by combining rocket data with the results based on older methods. Such a procedure is not justified and is likely to lead to unreliable results. For more reliable conclusion one must make use of rocket data from more than one station. The present calculation serve only to emphasise the desirability of carrying out rocket observations in different latitudes, particularly in the polar ranges.

If the temperatures for 30° N based as they are on rocket data, be taken as correct, and if the wind circulation is assumed to be due to the pressure gradient alone, then, from the observed wind velocities, one can compute the pressure distribution in the polar regions. Result of one set of such calculations is given in Table 4.

Table 4. *Height distribution of temperature in the polar region deduced from rocket data for latitude* 30° *N and the observed wind velocities*

Height	*Pressure (mb)*	*Temperature (°K)*
100 km	$4{\cdot}6 \times 10^{-5}$	230
120 km	$6{\cdot}0 \times 10^{-6}$	248
160 km	$1{\cdot}4 \times 10^{-7}$	342
200 km	$1{\cdot}6 \times 10^{-8}$	600

We note that the computed temperatures are in general lower than those shown in Fig. 1. The value 600° K for the 200 km level is in fair agreement with the estimate 750–850° K made by PETRIE from auroral spectra (1952).

6. CONCLUDING REMARKS

The data on which the computations of the wind velocities have been based are at best rough. Further, as already mentioned in Sec. 5 the computations do not take into account a number of disturbing factors which are likely to affect the wind system. It is, therefore, rather remarkable that notwithstanding these uncertainties, significant agreement has been obtained between the computed and the observed wind velocities. This shows that though a rigorous and comprehensive consideration of the dynamics of the high atmospheric regions may not be possible on account of the many difficulties involved, one may obtain reasonably good results by the adoption of simplified models based on plausible assumptions.

Acknowledgment—The author thanks Professor S. K. MITRA for suggesting the problem to him and for his helpful guidance in course of the preparation of the paper.

REFERENCES

BEYNON, W. J. G.	1948	Nature **162**, 887
CHAPMAN, J. H.	1953	Canadian J. Phys. **31**, 120
CHIU, WAN-CHENG	1951	Report on Upper Atmospheric Research, California Institute of Technology, March 31, 107
DEB, S.	1952	J. Atmos. Terr. Phys. **2**, 309
FENNEL, P.	1944	Bull. Amer. Met. Soc. **25**, 371
FERREL, O. P.	1948	Proc. I.R.E. **36**, 879
Geophysical Research Directorate, Record of Colloquium on upper atmosphere held in Harvard University	1950	July 31
GERSON, N. C.	1951	Trans. Amer. Geophys. Union **32**, 26
GERSON, N. C.	1952	Z. Angew. Phys. **4**, 81
GERSON, N. C.	1951	Rep. Prog. Phys. **14**, 316
HAVENS, R. J., KOLL, R. T. and LAGOW, H. E.	1952	J. Geophys. Res. **57** 59
HURUHATA, M.	1950	Rep. Ion. Res. Japan **4**, 137
HURUHATA, M.	1952	ibid **6**, 31
KELLOGG, W. W. and SCHILLING, G. F.	1951	J. Met. **8**, 222
MANNING, L. A., VILLARD, O. G., and PETERSON, A. M.	1950	Proc. I.R.E. **38**, 887
MATSUSHITA, S.	1950	Rep. Ion. Res. Japan **4**, 173
MAXWELL, A. and LITTLE, C. G.	1952	Nature **169**, 746
MEEK, J. H.	1949	Penn. State Coll. Rept., No. 5, A.M.C. Contract No. AF19 (122) 44
MITRA, S. N.	1949	Jour. I.E.E. **96**, 441
MITRA, S. K.	1952	The Upper Atmosphere, 2nd Ed., Asiatic Society, Calcutta
MOSSES, H. E. and TA-YOU-WU	1952	Phys. Rev. **87**, 628
MUNRO, G. H.	1950	Proc. Roy. Soc. **A202**, 208
PETRIE, W.	1952	Phys. Rev. **86**, 790

VOL. 4 1953

PHILLIPS, G. J.	1952	J. Atmos. Terr. Phys. **2**, 141
Report on Geophysical Discussion of the Royal Astronomical Society Observatory	1951	**71**, 104, Nature **167**, 625
Rocket Panel, The Harvard University	1952	Phys. Rev. **88**, 1027
RYLE, M. and HEWISH, A.	1950	Mon. Not. Roy. Astr. Soc. **110**, 381
SALZBERG, C. D. and GREENSTONE, R.	1951	J. Geophys. Res. **56**, 521
SHEPPARD, P. A.	1949	Science Progress **37**, 488
WHIPPLE, F. L.	1943	Rev. Mod. Phys. **15**, 246
YERG, D. G.	1951	J. Met. **8**, 244
YERG, D. G.	1952	J. Geophys. Res. **57**, 216
YUMURA, T.	1952	Rep. Ion. Res. Japan **4**, 173

OL.
4
953

Journal of Atmospheric and Terrestrial Physics, 1953 Vol. 4, pp. 44 to 52. Pergamon Press Ltd., London

Geomagnetic and ionospheric relationships

R. P. Waldo Lewis and D. H. McIntosh

Meteorological Office, Edinburgh.

(*Received* 17 *April* 1953)

ABSTRACT

The effect of magnetic "disturbance" on various parameters, both geomagnetic (measured at Eskdalemuir) and ionospheric (measured at Slough), is investigated by choosing days on which the K_p sum is markedly above or below those of adjacent days. Most of the parameters show a peak or trough corresponding to that of disturbance, and following it by periods of from 0 to 18 hours An extension of the investigation to measurements made at Huancayo shows mainly similar results.

Particular attention is called to some peculiar seasonal effects; also to features of the sporadic *E* layer occurrence at Slough and of the vertical movement of the *F*2 layer at Huancayo.

1. Introduction

Evidence is presented here on various aspects, other than those of diurnal variation, of the interrelations of geomagnetic and ionospheric phenomena.

2. Method

A choice was made of Greenwich days on which the sum of the 8 geomagnetic planetary three-hour-range indices K_p was a "peak" or a "trough" relative to adjacent days and on which the K_p sums were respectively above 25 and below 15. Attention was confined to "independent" occasions, and those days of marked maximum (or minimum) were included in preference to other days of perhaps greater (or smaller) K_p sum but differing little from neighbouring days. Totals of 103 peak and 83 trough days were selected from the K_p data for the period January 1947 to June 1952. The variations, on and around the chosen days, of various geomagnetic and ionospheric parameters were then calculated. Nearly all these have conspicuous diurnal variations the forms and amplitudes of which are systematically affected by varying magnetic disturbance. Since in this investigation it was desired to concentrate on effects of magnetic disturbance other than those of diurnal variation, the latter were largely removed by taking running means over a period of a day; the parameters were generally calculated at intervals of 4 hours so that the diurnal smoothing was achieved by taking running means of 6 values. The absence of a diurnal oscillation in the running means shows that a variation has been obtained which is free from effects of diurnal variation. It was found in practice that, with some of the parameters, some effect of diurnal variation survived this process of diurnal smoothing.

Owing to the appreciable secular changes of most of the parameters during the period considered, unequal representation of peak and trough occasions was found to give rise in some cases to differences in mean level quite unconnected with magnetic disturbance. In such cases the variations are shown as average departures (Δ) for a 24-hour period centred at each plotted point from the corresponding "all-day" mean level (differing slightly for the peak and trough occasions):

VOL.
4
1953

in the remaining cases the actual mean values, and not departures, are shown. Standard errors of selected plotted points are indicated. Results are discussed in the following sections.

3. Magnetic Variations

(a) *3-hour ranges*

Fig. 1 (a) represents the variation at Abinger of the amplitude of 3-hour ranges (R), in force units, calculated from the frequencies of the corresponding K indices.

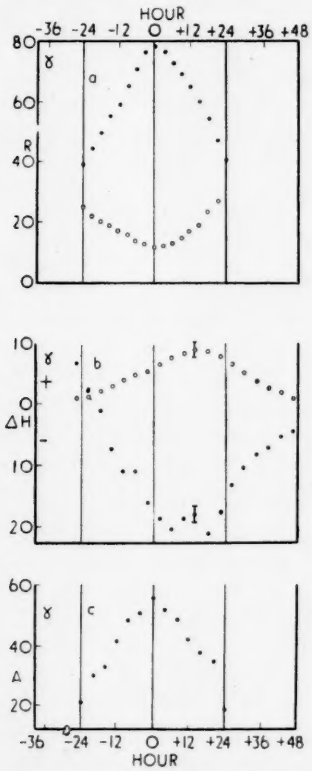


Fig. 1. Variation of (a) Abinger R (b) Eskdalemuir H (c) Eskdalemuir A round selected days of peak (· · ·) and trough (○ ○ ○) K_p sum, 1947–52.

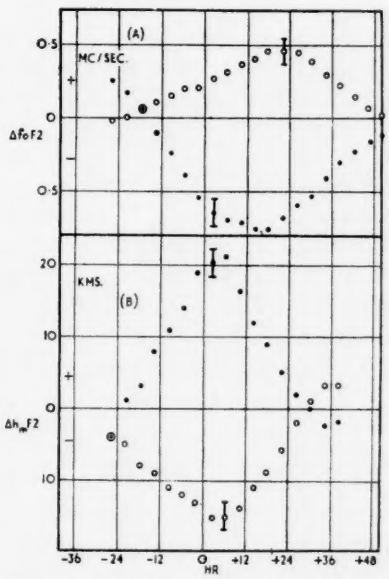


Fig. 2. Variation of (a) Slough Δf_0F2 and (b) Slough Δh_mF_02 round selected days of peak (· · ·) and trough (○ ○ ○) K_p sum, 1947–52.

This representation in absolute units is preferred to one of mean K index because of the non-linearity of the K index scale. Fig. 1 shows a fairly symmetrical R variation, with maximum or minimum occurring at hour "O" i.e. at 12 G.M.T. on the selected day; this type of variation is entirely due to the method of selection of the days, and has no connection with the ordinary diurnal variation of disturbance which has been removed by taking running means. The average peak day and trough day correspond to K sums of about 35 and 11 respectively.

(b) *H-force changes*

Fig. 1 (b) shows the corresponding H force variations at Eskdalemuir. (Values for Abinger were not available, but comparison over another period shows that at the two stations the changes of H force accompanying magnetic disturbance are statistically very close.) Fig. 1 (b) illustrates the known fact that the depression of H force associated with magnetic disturbance comes some time after the main disturbance as measured by range fluctuations. The time lag of H minimum behind R maximum is difficult to determine precisely because of an apparent double minimum in H, but appears to be about 18 hours. Examination shows that the double minimum in H force is probably caused by large changes in the amplitude of diurnal variation of H around the chosen peak days.

The falling-off of magnetic disturbance to minimum R is seen from Fig. 1 (b) to be associated with a rise of H force, lagging by about 18 hours. It is noteworthy that the ratio of the contrasting changes in H force is nearly the same as that of the corresponding changes in R.

(c) *Changes in amplitude of diurnal oscillation*

The changes in amplitude of daily variation of V force, and to a lesser degree of H force, accompanying magnetic disturbance provide parameters suitable for examining the variations of the S_D current system with which these changes are usually associated.

The mean V variations at Eskdalemuir round the peak days were determined at intervals of 2 hours and the amplitude (A) of daily variation of period 24 hours was calculated by the Fourier analysis of each successive 12 values. The results (Fig. 1 (c)), show that maximum A occurs at hour "0." The amplitude of daily variation of H force was also found to reach a maximum at this hour. These results strongly suggest that the S_D current system associated with disturbance reaches a maximum simultaneously with the range disturbance.

4. Ionospheric Variations

(a) *F2 layer critical frequency* (f_0F2)

The Slough f_0F2 variations round the peak and trough days are shown in Fig 2 (a). The expected depression of f_0F2 with increased disturbance can be seen and there is a corresponding increase of f_0F2 associated with the trough. In each case there is a lag of about 16 hours behind the associated magnetic range effects, a lag which cannot be considered to differ significantly from that of H force. The ratio (f_0F2 depression with increasing R)/(f_0F2 rise with decreasing R) is however definitely smaller than the corresponding ratio for the H force changes.

(b) *Heights and semi-thickness of F2 layer*

The Slough $F2$ layer variations in height of maximum ionisation density (h_mF2) are shown in Fig. 2 (b). h_mF2 rises and falls with R, lagging in either case by only about 4 hours and thus definitely preceding the f_0F2 changes. Simultaneous movements occur of the height of the lower edge of the $F2$ layer (h_0F2). These are in the same sense as the h_mF2 changes but the rise of h_0 is smaller than that of h_m

VOL.
4
195

by about 12 km and the fall is smaller by about 4 km. The assumption throughout of a parabolic distribution of electrons implies an expansion and contraction of the $F2$ layer by these amounts with increasing and decreasing magnetic disturbance respectively.

(c) *Sporadic E-layer (E_s) ionisation*

Fig. 3 (a) shows the variations of Slough fE_s. While an inverse relationship between fE_s and R is indicated by the peak curve the difference between the fE_s means in

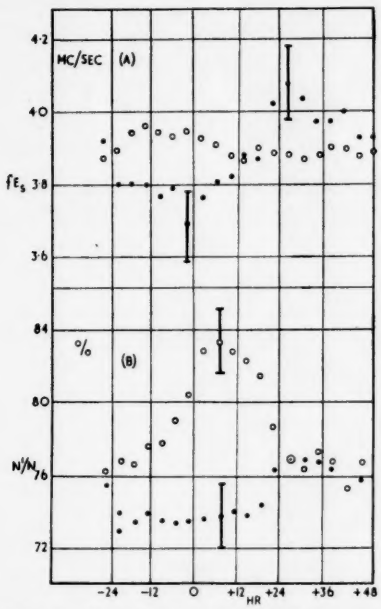


Fig. 3. Variation of (a) Slough fE_s and (b) percentage occurrence of Slough E_s (N'/N) round selected days of peak (• • •) and trough (○ ○ ○) K_p sum, 1947–52.

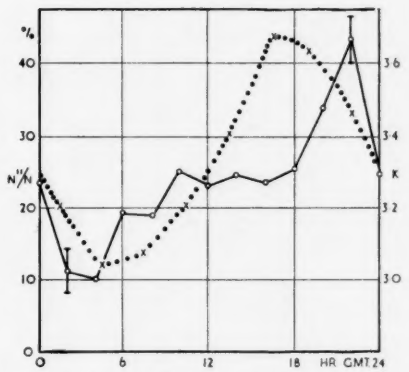


Fig. 4. Average diurnal variation of Abinger K index (× • • • ×) and Slough percentage non-occurrence (N''/N) of sporadic E (○——○) for years 1947–52.

the two groups of days is not proved significant from this data. In further examination of this point, however, we found that Slough fE_s was in all seasons significantly lower on international disturbed (D) days than on Q days by an average of about 0·2 Mc/sec.

Fig. 3 (b) shows the rather unexpected result that the percentage frequency (N'/N) of E_s occurrence is greater at times of magnetic quiet than at disturbed times. (The only relation so far confirmed between the occurrence of E_s ionisation and magnetic disturbance is a direct relation at high latitudes (Appleton, 1936).) The variations in frequency of occurrence appear to lag about 4 hours behind the corresponding variations of R i.e. about the same as the lags in the vertical movements of the $F2$ layer.

Another aspect of this suggested inverse relation between geomagnetic disturbance and E_s occurrence is shown in Fig 4 where average features of diurnal

variation are compared. Here percentage *non*-occurrence (N''/N) of E_s is plotted for easier comparision. The general similarity of the two curves lends some support to this suggestion of a relationship between them, but it is apparent that there is an important influence other than geomagnetic disturbance in the occurrence of Slough E_s. Comparison of seasonal features of the two phenomena leads to a similar conclusion: the winter, equinox, summer values of Slough N''/N, meaned over all hours in the years 1947–52, are respectively 22, 31, 17 per cent; non-occurrence of E_s thus resembles disturbance in having a clear equinoctial maximum, but differs from it in having a higher value in winter than in summer. Meek (1949) has distinguished five types of E_s ionisation in high latitudes. It seems likely that more than one type is recorded also at Slough, and that only a fraction of the recorded occasions is sensitive to geomagnetic changes.

5. Seasonal Variations

The seasonal variations of selected parameters are shown in Fig. 5. The method of selection of the days implies compensating changes in level of magnetic disturbance

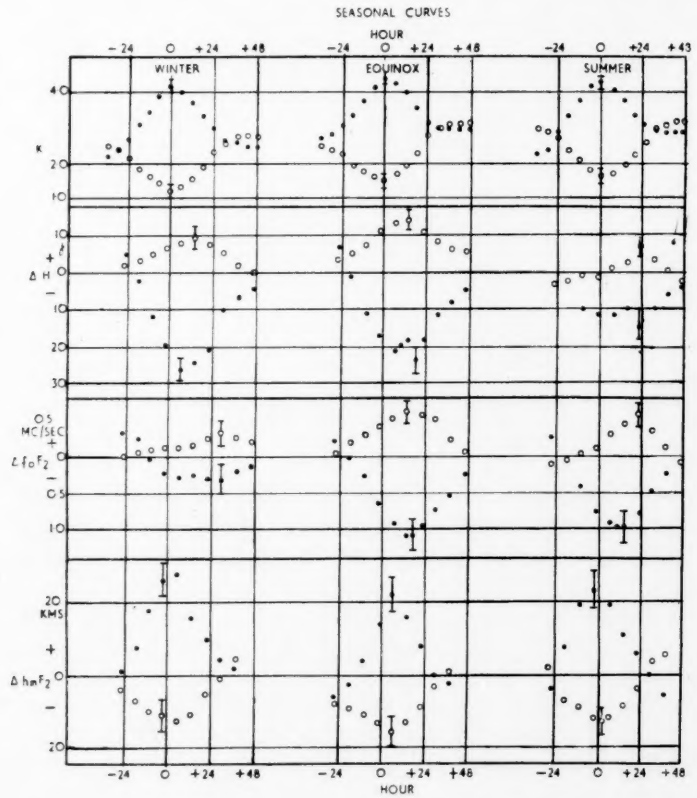


Fig. 5. Average winter, equinox, summer variations of (a) Abinger K (b) Eskdalemuir ΔH (c) Slough Δf_0F2 and (d) Slough Δh_mF2 round selected days of peak (· · ·) and trough (○ ○ ○) K_p sum, 1947–52.

some time before and after the average selected days. The resulting intersection of peak and trough K curves is shown in Fig. 5, with corresponding intersections occurring somewhat later in h_mF2 and still later in H force and f_0F2.

A small degree of asymmetry may be noted in the average form of magnetic disturbance at Abinger, the fall of K following the peak occurring more slowly than its previous rise. Similar asymmetry may be noted in the other elements shown in Fig. 5.

Despite the close association of the elements in Fig 5, there are important contrasts in their seasonal characteristics: h_mF2 changes insignificantly with season while Δf_0F2 and ΔH have definite minima in winter and summer respectively. We have tested the reality of these features by independent methods.

6. Non-cyclic Changes

It is well known that there is a negative non-cyclic change (N.C.C.) of H force on the average magnetically disturbed day. The time lag of Slough f_0F2 and h_mF2 behind magnetic disturbance gives rise to N.C.Cs. (respectively negative and positive) in these also. The variation with time of the 24-hour N.C.Cs. round the peak days is shown in Fig. 6. As would be expected, there is practically no N.C.C. of K index from beginning to end of the selected peak day, i.e. $K_{+12} - K_{-12} \simeq 0$: the small positive value is readily explained by the slightly asymmetrical form of the disturbance curve. Assuming similar asymmetry in the other elements, their lag behind the magnetic range disturbance is given by the delay with which zero N.C.C. occurs, i.e. approximately 4 hours for h_mF2 and 14 hours for f_0F2 (Fig. 6). Determination by this method of the lag of H force at Eskdalemuir is complicated by the large disturbance diurnal effects.

7. Further Geomagnetic Connections

The statistical relations of magnetic disturbance with various other parameters were investigated in less detail. Results are as follows:

(a) *Relative sunspot number*

The mean relative sunspot number was found to be above average on each of the 6 days round the chosen peak occasions and very slightly below average round the trough days. The mean difference of about 11 units of sunspot number between the two groups of days was however barely significant in terms of its general variability, while the mean day-to-day variations in the two individual groups were quite insignificant. Comparison of this result with the very significant variations already described leads to the inference that the geomagnetic and ionospheric phenomena are more closely related than is either with the solar activity as measured by sunspot number.

(b) *Ionospheric absorption*

The Slough ionospheric absorption measurements were examined for evidence of association with the geomagnetic and ionospheric peaks and troughs.

The mean values of the absorption coefficient A and of the absorption index $-\log_e \rho$ measured at a frequency of 2·0 Mc/sec. varied little in both groups between

days −3 and +3, and no significance could be attached to the differences between corresponding values in the two contrasting groups.

(c) *E layer critical frequency*

Departures of Slough E layer critical frequency ($\Delta f_0 E$) from the appropriate monthly means were examined in relation to the selected peak and trough days. Owing to the absence of measurements for all night and twilight hours and to the

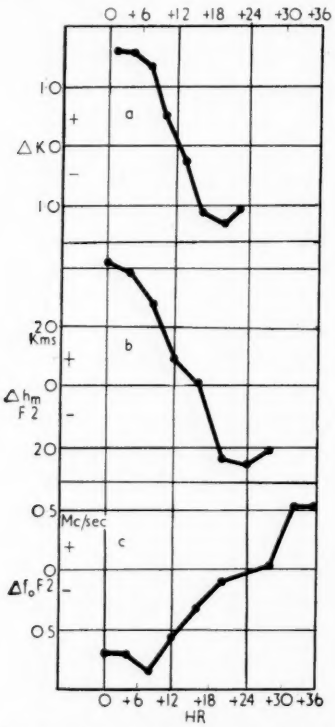


Fig. 6. Variation of 24-hour N.C.Cs. in (a) Abinger K, (b) Slough $h_m F_2$, and (c) Slough $f_0 F_2$, round selected days of peak K_p sum, 1947–52. The N.C.Cs. are over periods ending at the times shown; hour "0" indicates middle of peak day.

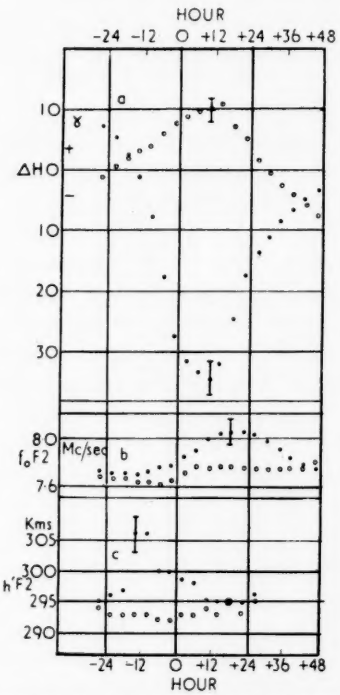


Fig. 7. Variation of (a) ΔH (b) $f_0 F2$ and (c) $h'F2$ at Huancayo round selected days of peak (• • •) and trough (○ ○ ○) K_p sum, 1940–43.

general irregularity of the effects any attempt at timing was impossible, but evidence was obtained that $f_0 E$ on a peak day and the day following was significantly lower by about 0·4 Mc/s than it was on a trough day and the day following.

8. Huancayo Variations

Effects near the equator analogous to some of those discussed above are shown in Fig. 7 in which the variations at Huancayo of H force, $f_0 F2$ and $h'F2$ (minimum virtual height) are represented. The variations are those for 50 selected K_p peak

VOL. 4 1953

and 37 K_p trough days from 1940–44 and are shown on the same scale as in Figs. 1 and 2. The Huancayo K variations are not reproduced but resemble those at Abinger in having maximum and minimum, corresponding to peak and trough occasions respectively, at 12 G.M.T. The Huancayo H changes are bigger than those at Eskdalemuir and, lagging behind R by only about 10 hours, probably precede the corresponding Eskdalemuir variations by some hours. The Huancayo $F2$ layer changes of f_0 and h' are, on the other hand, much less than at Slough: f_0, showing the known positive equatorial variation with increasing disturbance,

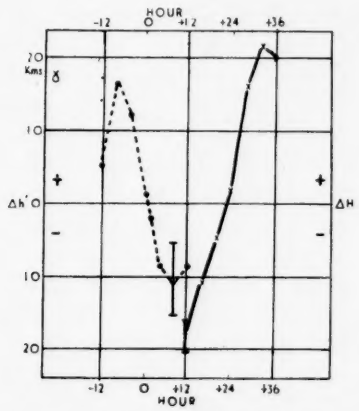


Fig. 8. Variation of 24-hour N.C.Cs. in Huancayo $h'F2$ (· · · · · ·) and H force (×—×) round selected days of peak magnetic disturbance 1938–39 and 1944–45.

may possibly be delayed by a few hours compared with the H changes; the h' changes are in the same sense as the corresponding Slough changes but occur relatively even earlier. Fig. 7 strongly suggests, in fact, that the $h'F2$ reactions at Huancayo to magnetic disturbance reach their maximum some 12 hours *before* the ranges, as compared with a *delay* of about 4 hours at Slough. The result of a test with independent Huancayo material of this time relationship is shown in Fig. 8, in which the time variation of the N.C.Cs. in $h'F2$ and H force may be seen to agree closely with the results of Fig. 7.

The variations of Huancayo E layer critical frequency (fE) associated with increasing and decreasing magnetic disturbance were found to be insignificant.

9. Discussion

The main method of analysis in this paper differs little from a statistical examination of various parameters on magnetically disturbed, as opposed to quiet days; concentration on independent occasions of marked peaks and troughs and the use of the "superposed-epoch" method have however enabled us to compare sharply opposed effects and to assess the average time relations of various elements. We have previously (Lewis and McIntosh 1953) called attention to the very closely related "storm-time" variations that occur at Slough-Eskdalemuir in f_0F2 and H force and suggested a possible slight precedence of the magnetic by

the ionospheric changes. Any time difference between these elements has however proved too small to be determined in the material considered here.

Comparison of the Slough-Eskdalemuir results with those for Huancayo makes it apparent that the related $F2$ layer and H force changes are not cause and effect in the sense that the expanding $F2$ layer with varying electron density reacts directly on the magnetic field measured at the ground: other considerations show, in any case, that any such direct effect must be very small. The relationships between the phenomena appear to us to be not inconsistent with MARTYN's (1951) view that the measured ionospheric reactions are the result of vertical electronic drifts induced by the Chapman-Ferraro ring current, but there are obviously many compelxities and anomalies yet to be explained. Examples of these are the existence at Slough of an inverse relation between E_s occurrence and geomagnetic disturbance as opposed to a direct relationship at higher latitudes (§ 4 (c)); the contrasts in seasonal characteristics of closely related phenomena (§ 5); the large differences in time of maximum displacement of the $F2$ layer at Slough and Huancayo (§ 8). A possible explanation of the early movement of the $F2$ layer at Huancayo is that the movement is short-lived and is confined to a time at or near occasions of "sudden commencement" of storms. There is support for this in that the statistical time lag of magnetic range maxima behind sudden commencements is about 12 hours (the same as that between $h'F2$ and R), but we are uable to confirm or refute this possible explanation from available data. Whatever the reason for the Huancayo $h'Fs$ effect may be, it can hardly also serve to explain the corresponding effect noted at Slough.

MARTYN (1953) has recently stated that he finds no evidence for the view that ionospheric storms travel in a matter of hours or days, from high to low latitudes. Our results support MARTYN's statement in that we find that corresponding f_0F2 changes occur simultaneously in high and low latitudes (as do the related magnetic R changes). Such differences as have appeared in the times of corresponding phenomena have in fact been in the opposite sense e.g. the vertical movements of the $F2$ layer and probably also the maximum depression of magnetic H force appear to occur earlier at Huancayo than at Slough and Eskdalemuir respectively.

VOL. 4 1953

REFERENCES

APPLETON, E. V., NAISMITH, R., and INGRAM, L. J.	1936	Phil. Trans. Roy. Soc. A, **236**, 191
MEEK, J. H.	1949	J. Geophys. Res. **54**, 339
LEWIS, R. P. WALDO and McINTOSH, D. H.	1953	J. Atmosph. Terr. Phys.
MARTYN, D. F.	1951	Nature **167**, 92
MARTYN, D. F.	1953	Nature **171**, 14

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 53 to 67. Pergamon Press Ltd., London

World-wide diurnal variations in the $F2$ region

C. W. Allen

University College, London

(*Received* 25 *February* 1953)

ABSTRACT

A study has been made of the diurnal variations of $F2$ critical frequencies and virtual heights as a function of phase of sunspot cycle, season, and geographic and geomagnetic position. The main results are derived from ten stations that are well observed at both sunspot maximum and minimum.

The chief characteristics of f^0F2 may be described in terms of (a) the sunspot minimum anomaly, (b) the sunrise anomaly, and (c) the diurnal range anomaly. The features of these anomalies are evaluated quantitatively and then, as far as possible, the anomalies are extracted from the data to determine the normal curves. The world distributions of indices representing the anomalies are studied. The variations of the $F2$ virtual height are not related in any very significant manner to either the sunspot cycle or the f^0F2 diurnal curves.

Some attempt at explanation of the anomalies has been made but no complete theory is available. Some effects are thought to be thermal or tidal oscillations, and evidence is given for the conclusion that $F2$ electrons disappear by a decay law that is independent of gas pressure.

1. Introduction

It is well known that the $F2$ region of the ionosphere is anomalous as compared with the regular E and $F1$ regions. The observed characteristics show anomalies in the diurnal variations, day-to-day variations, seasonal variations, solar cycle variations, and variations with geographic and geomagnetic positions. The $F2$ layer is very sensitive to external disturbances, but the short period variations may be averaged out into monthly data which are then fairly regular. After the sporadic variations have been eliminated there remain many persistent features which do not agree with any known theory and call for an explanation. They require a systematic analysis in order to present them in a quantitative form.

There are probably several superimposed anomalies, and it may be possible to study them one by one. If any anomaly can be evaluated either physically or empirically one should be able to extract its effect from the data and proceed with the corrected data to study the next anomaly or the physical principles. The procedure provides quantitative information for a physical study of both the anomaly and the corrected results.

The analysis in the present paper is an attempt to use this outlook. Certain regularities in the diurnal variation of critical frequencies f^0F2, were noted earlier [1] and gave promise of leading to a result that would be systematic on a world-wide scale. This regular anomaly has been extracted and the corrected data examined. The analysis is not influenced by theoretical considerations except that a Chapman layer [2] is regarded as the ideal of normality and departures from it are called anomalies.

2. Diurnal Variations of Ten Stations

The data analysed in detail are from ten widely distributed stations for which ionospheric characteristics at sp-max (sp = sunspot) and sp-min are well observed. All information was extracted from the "Ionospheric Data" publication of the

Central Radio Propagation Laboratory, Washington. The stations used are shown in Table 1.

Table 1

Station	*Geographic* *Lat.*	*Long.*	*Magnetic* *Lat.*	*Long.*	*Time standard*
Fairbanks, Alaska	64°·9 N	147°·8 W	64°·7 N	258°·2 E	150° W
Washington, U.S.A.	38°·7 N	77°·1 W	50°·3 N	350°·5 E	75° W
San Francisco, U.S.A.	37°·4 N	122°·2 W	43°·6 N	298°·8 E	120° W
Baton Rouge, U.S.A.	30°·5 N	91°·2 W	41°·0 N	334°·8 E	90° W
San Juan, Puerto Rico	18°·4 N	66°·1 W	29°·9 N	3°·9 E	60° W
Maui, Hawaii	20°·8 N	156°·5 W	20°·9 N	268°·5 E	150° W
Huancayo, Peru	12°·0 S	75°·3 W	0°·6 S	353°·7 E	75° W
Brisbane, Australia	27°·5 S	153°·0 E	35°·7 S	227°·0 E	150° E
Watheroo, Australia	30°·3 S	115°·9 E	41°·7 S	185°·9 E	120° E
Canberra, Australia	35°·3 S	149°·0 E	44°·0 S	225°·0 E	150° E

Various diurnal curves are shown in Figs. 1 and 2. For each station the top section is averaged from the four months of the northern solstice May, June, July, Aug., the middle section for the equinoxial months Mar., Apr., Sept., Oct., and the lower section for the southern solstice Nov., Dec., Jan., Feb. Within each section the top continuous curve is the diurnal variation of f^0F2 for sunspot number $R = 100$, and the superimposed broken line curve is the same with the sp-min anomaly (of § 3) eliminated. The next continuous curve (approximately parallel) shows f^0F2 for $R = 0$. As explained in the earlier paper [1] these results are obtained by extrapolating and interpolating among the curves for the individual years of sp-max and sp-min. The f^0F2 scale for these curves is logarithmic and hence when curves are parallel the variations are proportional. However, the lack of parallelism shows some systematic trends that are demonstrated in the diagram by the next curve which is simply the difference between the $R = 100$ and $R = 0$ curves. The difference curves are discussed in § 3. The noon values of f^0F1 at $R = 100$ and $R = 0$ are also shown in Figs. 1. and 2.

Below the dividing line in each section are the curves of the virtual heights $h'F1$ and $h'F2$ on a linear scale. The sp-max curves are on top and the curves are separated by 200 km to avoid superposition. The 250 km level is inserted in each case. The $h'F2$ values showed a much less systematic variation with the sp-cycle than f^0F2, and consequently it has been thought sufficient to represent the curves at sp-max and sp-min without attempting to deduce curves for precisely $R = 100$ and $R = 0$. $h'F1$ results are included as broken lines.

3. Sunspot Minimum Anomaly

The most striking anomalies in the f^0F2 curves of Figs. 1 and 2 are at sp-min. Our aim will be to study this sp-min anomaly quantitatively in this section, and later study the effects of its removal.

The anomalies tend to be opposite in summer and winter. In summer—well illustrated by Washington northern solstice—the sp-min f^0F2 curves are characterised by a maximum near 20 h, a rapid night decline, and deep presunrise

VOL. 4 1953

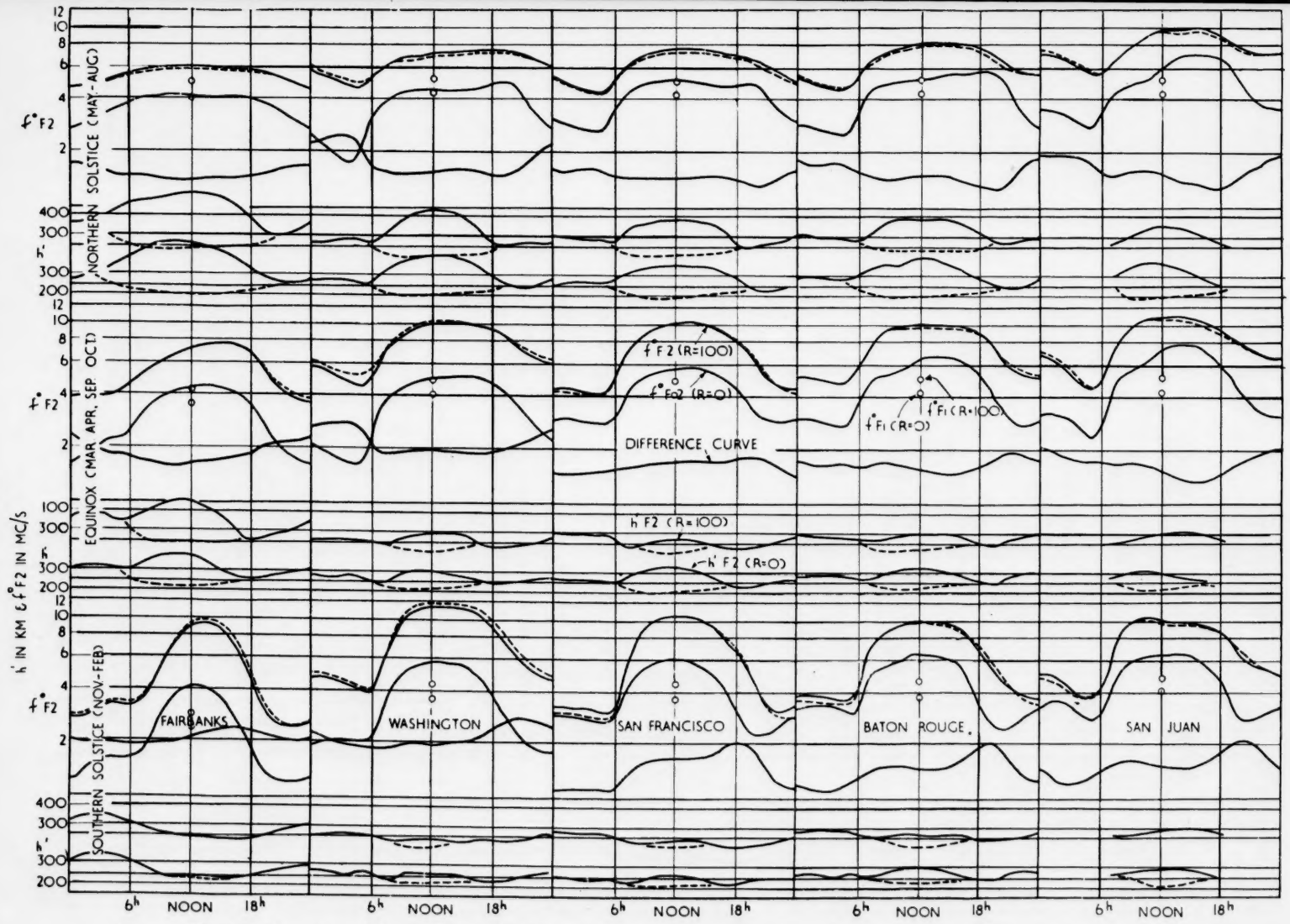


Fig. 1. Diurnal variations of f^0F2 and $h'F2$ at sp-max ($R = 100$) and sp-min ($R = 0$); also the difference curves.
..... = f^0F2 ($R = 100$) corrected for sp-min anomaly; also $h'F1$.
○ = noon f^0F1 ($R = 100$ on top and $R = 0$ on bottom).

minimum. In winter—well illustrated by Watheroo northern solstice—the spmin f^0F2 curves have opposite characteristics, a minimum near 20 h, a rise during the night, and a presunrise maximum. Most of the diurnal variations show some evidence of these characteristics in either the summer or the winter form, but the characteristics disappear at some intervening season.

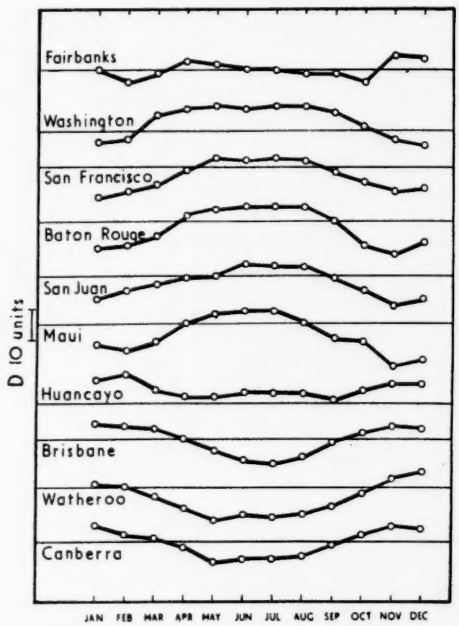


Fig. 3. Seasonal variation of D.

Table 2. Measurements of index D representing sunspot minimum anomaly

Station	*Jan*	*Feb*	*Mar*	*Apr*	*May*	*Jun*	*Jul*	*Aug*	*Sep*	*Oct*	*Nov*	*Dec*	*N Sols*	*Equinox*	*S Sols*	*Range N–S*	*Year*
Fairbanks	0	−4	−1	+3	+2	0	0	−1	−1	−4	+5	+4	0	−1	+1	−1	0
Washington	−4	−3	+5	+7	+8	+7	+8	+8	+6	+1	−3	−5	+8	+5	−4	+12	+3
San Francisco	−10	−8	−6	−1	+3	+2	+3	+2	−2	−5	−8	−7	+3	−4	−9	+12	−3
Baton Rouge	−9	−8	−5	+2	+4	+5	+5	+5	0	−8	−11	−7	+5	−3	−9	+14	−2
San Juan	−8	−5	−3	−1	0	+4	+3	+3	−1	−5	−10	−8	+3	−3	−9	+12	−3
Maui	−7	−9	−6	0	+3	+4	+4	+0	−5	−6	−14	−12	+3	−4	−11	+14	−4
Huancayo	+7	+9	+4	+2	+2	+3	+3	+3	+1	+4	+6	+6	+3	+3	+7	−4	+4
Brisbane	+5	+4	+3	0	−4	−7	−8	−6	−1	+2	+4	+3	−6	+1	+4	−10	0
Watheroo	+1	0	−3	−7	−11	−9	−10	−9	−6	−2	+3	+5	−10	−4	+2	−12	−4
Canberra	+5	+2	+1	−2	−7	−6	−6	−5	−1	+2	+5	+4	−6	0	+4	−10	−1

As the anomaly is consistently more severe at $R = 0$ than $R = 100$ it may be measured quantitatively by means of the difference between these two curves. The difference curves show the characteristics of the anomaly rather clearly, and show in particular that we can measure it by one index with positive and negative values. To make such a measurement we proceed, somewhat as in a harmonic analysis, to measure ordinates at the expected maxima and subtract ordinates at the expected minima. Our difference curve index D was obtained by adding

VOL.
4
195

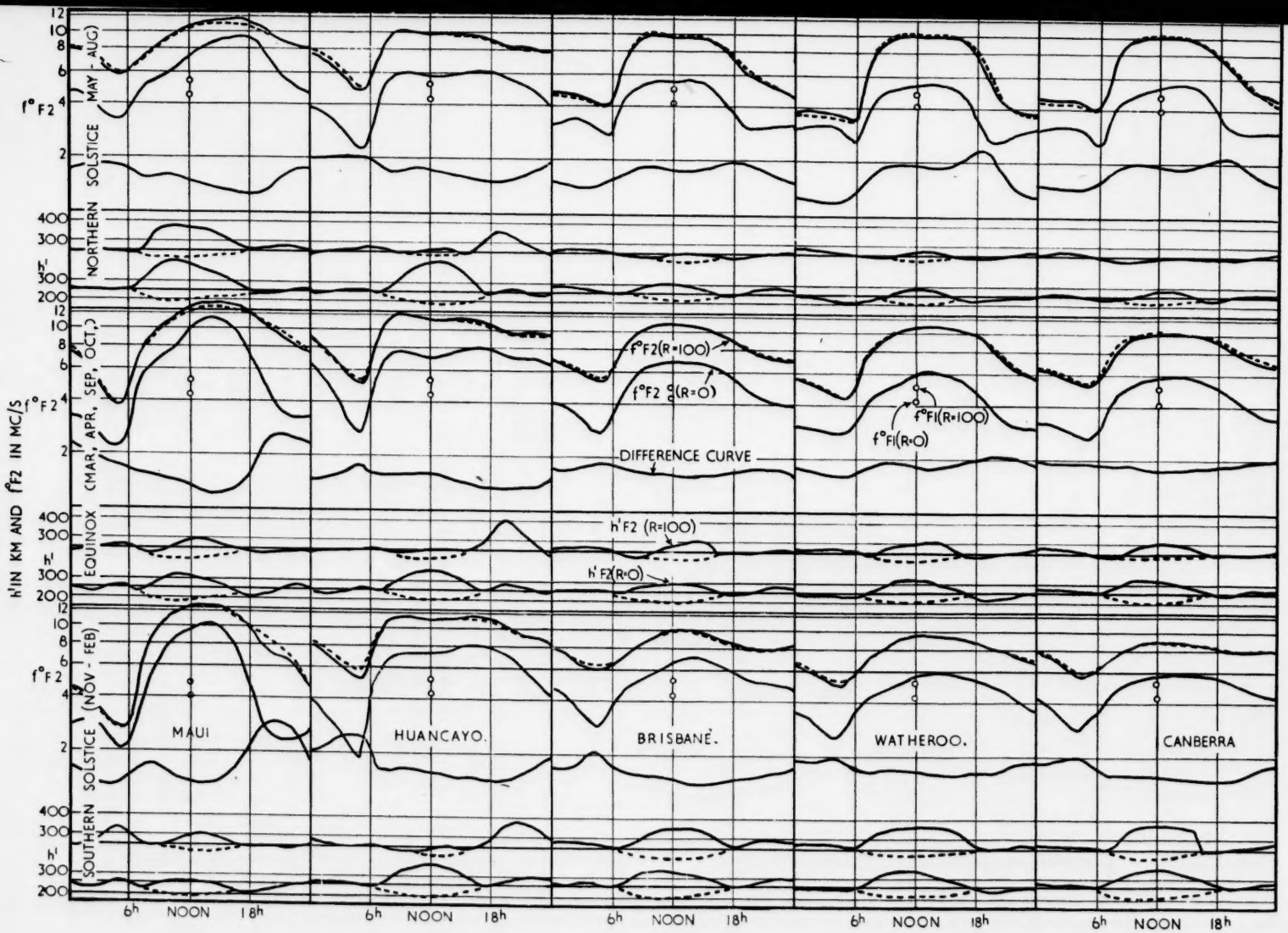


Fig. 2. Diurnal variations of f^0F2 and $h'F2$ at sp-max ($R = 100$) and sp-min ($R = 0$); also the difference curves.
..... = f^0F2 ($R = 100$) corrected for sp-min anomaly; also $h'F1$.
○ = noon f^0F1($R = 100$ on top and $R = 0$ on bottom).

ordinates at 4 h and 15 h and subtracting those at 9 h and 20 h. Summer type anomalies have positive values, and the scale of the index is such that for $D = +10$ and -10 the logarithmic difference curves are given in Fig. 5. With the exception of the arctic station at Fairbanks, where $D \simeq 0$, all stations vary in the sense that $D(\text{summer}) > D(\text{winter})$.

The monthly curves have been examined to obtain the run of D throughout the year, and these values are shown in Table 2 and Fig. 3. The annual variation are in the form of cosine waves with no appreciable change of phase. Hence the characteristics of the annual D variation may be represented by an amplitude and a mean value. The last two columns of Table 2, the range of D, i.e. the

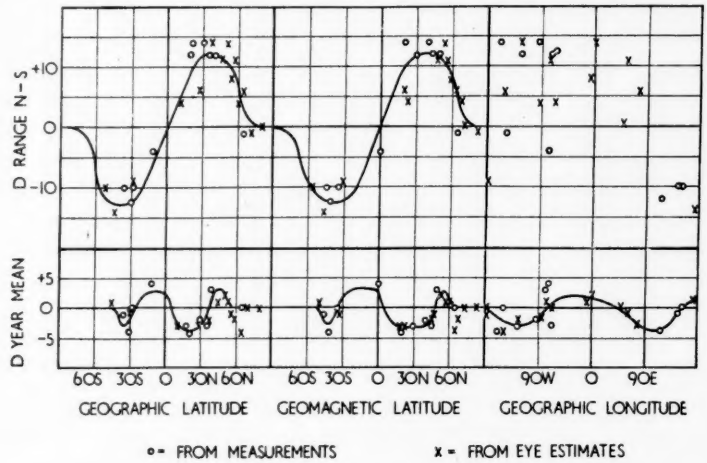


Fig. 4. Variation of D with latitude and longitude.

northern solstice minus southern solstice value, and the yearly mean D, give a suitable representation. They have been plotted against geographic and geomagnetic latitude and geographic longitude in Fig. 4. The results for the ten stations studied in detail have been supplemented by a visual examination of the sp-min f^0F2 curves of 20 other stations and the results given in Table 3. These eye estimates are shown as crosses in Fig. 4.

It is seen that the range of D changes fairly regularly from zero in the arctic to a constant value at mid latitudes, and a change of sign at the equator. A representative symmetrical curve is fitted to the values in Fig. 4, and fits geographic and geomagnetic latitudes equally well. No systematic change of the range in D with longitude is found or expected.

The mean of D is a decided characteristic of the diurnal f^0F2 variation; for example, it represents the main difference between the Washington and San Francisco curves. Nevertheless there is no very apparent regularity in this value which is also plotted in Fig. 4. There is some tendency for the values to follow the curves drawn, but these are not symmetrical with latitude or longitude and are quite uncertain. However, the reality of the variations in mean D is assured by the similarity of results for near stations such as Washington and Ottawa, San

VOL. 4 1953

Table 3. Eye estimates of D from diurnal f^0F2 curves

Station	Geographic Lat.	Geographic Long.	Geomag. Lat.	D N Sole.	D Equi-nox	D S Sols	D Range N–S	D Year mean
Burghead	55°·7 N	3°·5 W	+58°·9	+5	+3	−3	+8	+1
Great Baddow	51°·7 N	0°·5 E	+54°·4	+9	+3	−5	+14	+2
Ottawa	45°·4 N	75°·7 W	+56°·7	+6	+3	−5	+11	+1
Stanford	37°·4 N	122°·2 W	+43°·6	+5		−8	+14	−2
Churchill	58°·8 N	94°·3 W	+68°·5	0	−1	−4	+4	−2
College	64°·9 N	147°·8 W	+64°·4	−1	−3	−7	+6	−4
Delhi	28°·6 N	77°·2 E	+19°·0	0	−4	−6	+6	−3
Trinidad	10°·6 N	61°·3 W	+21°·8	−1	−3	−5	+4	−3
Reykjavik	64°·1 N	21°·7 W	+70°·0	+1	+1			
Baffin Is.	70°·5 N	68°·6 W	+82°·3	0	0	+1	−1	0
Sverdlovsk	56°·8 N	61°·1 E	+48°·3	+4	0	−7	+11	−1
Madras	13°·0 N	80°·2 E	+3°·1	+8	+5			
Tykhi Bay	80°·3 N	52°·8 E	+71°·4	0	0	0	0	0
Moskow	55°·8 N	37°·6 E	+51°·1	+7	+1			
Christmas Is.	2°·0 N	157°·0 W	+2°·4		+5	+8		
Kermadec Is.	29°·2 S	177°·9 W	−33°·5	−6	0	+3	−9	−1
Christchurch	43°·5 S	172°·6 E	−48°·0	−6	+2	+8	−14	+1
Campbell Is.	52°·0 S	169°·0 E	−56°·7	−4		+6	−10	
Simontown	33°·9 S	18°·7 E	−32°·7			+4		
Cape York	11°·0 S	142°·4 E	−20°·8			+4		

Francisco and Stanford, Trinidad and San Juan, Great Baddow and Burghead, Canberra and Brisbane. Apparently the variation of mean D forms no simple world-wide pattern, and is too complicated to be revealed by the 30 stations available at the last sp-min.

Fig. 5 shows mean difference curves for stations with well developed positive (summer) and negative (winter) characteristics. The scale of the variations is adjusted to represent $D = +10$ and -10 on our scale. The positive and negative difference curves are almost mirror images. Both have maxima or minima at 04 h and 19 h 30 m local time and a suggestion of secondary maxima and minima at 9 h and 15 h. These curves are dominated by stations at middle latitudes but the tropical station at Huancayo is of similar positive type. On the other hand the Maui curves are abnormal and show that the world picture of the sunspot minimum anomaly is not quite as simple as implied.

The level of the difference curve when there is no anomaly ($D = 0$) represents the main sp-cycle variation of f^0F2. The mean log height is 0·23, giving an f^0F2, $R = 100/R = 0$ ratio of 1·70. The variations from this mean value appear to be random.

4. Corrected Sunspot Maximum Curves

There is no accurate means of determining the amount of sp-min anomaly retained in the f^0F2 curves at sp-max. However, the characteristic peaks and hollows near 4 h and 20 h have been sought in the diurnal curves and it is estimated that

there is about 20% as much of this type of variation in the sp-$R = 100$ logarithmic f^0F2 curves as in those for sp-$R = 0$. This small amount of variation may be eliminated by subtracting 25% of the variable part of the difference curves. The result of this correction is shown in Figs. 1 and 2 where it is seen that the changes introduced by it are quite unimportant.

A normal diurnal f^0 curve controlled by solar ionisation and steady electron recombination would have a smooth maximum between noon and 18 h and would descend smoothly to a single minimum near sunrise. It might be expected to have approximately the same range (logarithmically) in summer and winter. There appear to be two main departures from this ideal which we might call (a) the sunrise anomaly, and (b) the diurnal range anomaly.

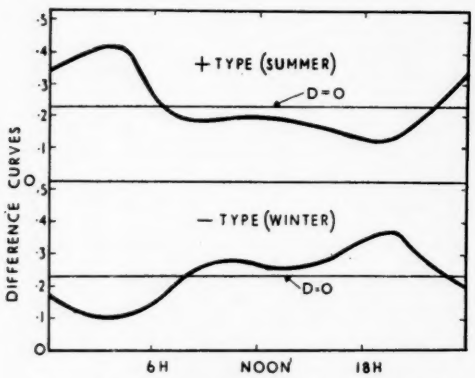


Fig. 5. Difference curves for $D = +10$ and $D = -10$. (Ordinates are in $\log_{10}$)

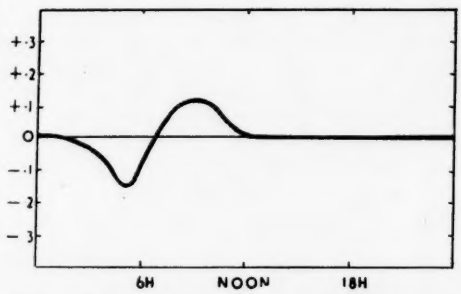


Fig. 6. Log departure from normal f^0F2 curve for $S = 5$ and ground sunrise at 6 h.

(a) There is a widespread characteristic wave in f^0F2 near sunrise. About one hour before ground sunrise there is a rather sharp minimum followed about 4 hours later by a maximum. The logarithmic departure from the expected curve for a 6 h sunrise is shown in Fig. 6. Eye estimates S of the magnitude of this effect have been made from Figs. 1 and 2 and from a number of sp-max diurnal curves of other stations near the equator. The scale is such that the curve of Fig. 6 has the value $S = 5$. The values of S are given in Table 4 and plotted against latitude in Fig. 7. The points are fitted to symmetrical mean curves and it is seen that S is small in summer, large in winter, and large throughout the year near the equator. The effect never appears to be negative. It follows geomagnetic latitude somewhat better than geographic.

(b) When the sunrise anomaly has been extracted there is still much change in the magnitude of the diurnal variation which is small in summer, large in winter, and small at the equator. A measure V of the diurnal variation is given by the log ratio max/min measured on the f^0F2 diurnal curves from which S has been extracted. The values of V are also given in Table 4 and fitted to symmetrical curves in Fig. 7.

The sp-$R = 100$ logarithmic diurnal curve for any latitude and season could be reconstructed from S and V in the following manner. Draw a curve with a broad

VOL. 4 1953

maximum at 15 h, a smooth decline to a minimum near $F2$ sunrise at which time f^0 is half its maximum value, and then a smooth rise to the maximum. On to this superimpose a curve $\frac{1}{2}(0{\cdot}3 - V)\cos 15t^0$ where t is the time in hours and 0·3 is taken as the mean value of V, then superimpose the appropriate correction for S using Fig. 6. Curves constructed in this way will bear the main characteristics of the f^0 diurnal variation. For other values of R one could assume a linear relation

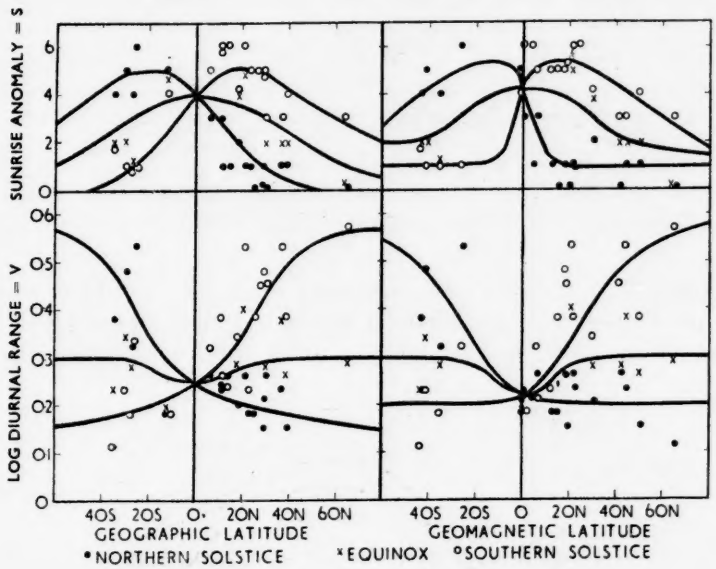


Fig. 7. Variation of S and V with latitude. The curves are smoothed from the individual values.

between f^0 and R and use the difference curve index D to determine the magnitude of the variation.

Any study of the variation of f^0F2 with locality and season must face the difficulty of selecting the appropriate time of day for the measurements. It appears that when the diurnal range is great the noon values are abnormally high, and vice versa. We should obtain more representative values of f^0F2 if the variations in diurnal range were extracted. From log f^0F2 at diurnal maximum we therefore subtract $\frac{1}{2}(V - 0{\cdot}3)$ to obtain the diurnal maximum corrected for variations in the diurnal range. Values have been obtained from Figs. 1 and 2 and are supplemented by measurements from a number of tropical stations in the years 1948–49. These latter values are decreased 15% to bring them to $R = 100$. Values are given in Table 4 and plotted in Fig. 8. Summer values tend to be greater than winter but the difference is not large nor very systematic. The change of critical frequency with latitude is fairly clear and seems to follow geomagnetic latitude somewhat better than geographic. The relation between equinoxial noon f^0F2 and geomagnetic latitude reported by APPLETON [3] will be seen to agree with the results of Fig. 8 when combined with the V range variations of Fig. 7.

As the corrections S and V are not based on any physical reasoning it may be

useful to determine mean diurnal f^0F2 curves without extracting these factors. In equatorial regions the curves are too complicated to justify making a mean but

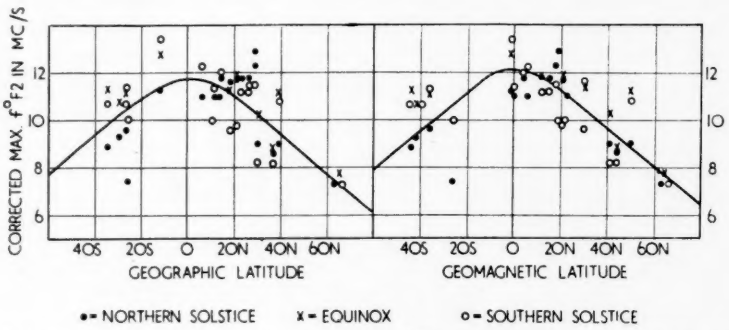


Fig. 8. Variation with latitude of daily maximum f^0F2 corrected for variations of S and V, at $R = 100$.

in mid latitudes they run to a fairly regular pattern. Fig. 9 shows mean curves which should represent a hypothetical station at geographic latitude 35° or geomagnetic latitude 40°, sp-$R = 100$, and with only the small sp-min anomaly extracted.

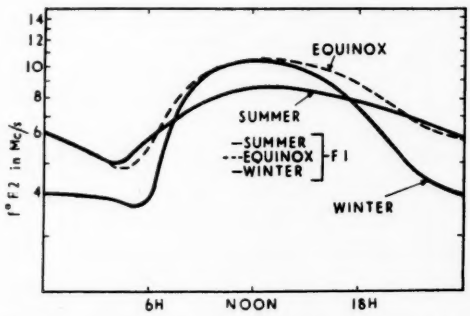


Fig. 9. Diurnal variation of f^0F2 at sp-max for mid latitude stations.

5. VARIATIONS IN $h'F2$

The curves for the diurnal variation of the virtual height $h'F2$ set out in Figs. 1 and 2 do not reveal any clearly systematic differences between sp-max and sp-min. From the diagrams we find the following height statistics.

Cases when $h'F$ was higher	*F1 noon*	*F midnight*	*F2 noon*
at sp-max than sp-min	20	14	9
at sp-min than sp-max	6	9	19

When we consider that group retardation by $F1$ will tend to raise the measured noon values of $h'F2$ at sp-min we may conclude that there is a slight general tendency towards greater actual heights at sp-max than at sp-min.

VOL. 4 1953

Table 4. Anomalies S and V, and f^0F2 at sp-R = 100, N = Northern Solstice, Eq = Equinox, S = Southern Solstice

Station	Geograph. Lat.	Geograph. Long.	Geomag. Lat.	S N	S Eq	S S	V N	V Eq	V S	Corrected Max. f^0F2 (R = 100) N	Corrected Max. f^0F2 (R = 100) Eq	Corrected Max. f^0F2 (R = 100) S
							in logs			Mc/S		
Fairbanks	65 N	148 W	+65	0	0	3	0·11	0·29	0·57	7·3	7·8	7·3
Washington	39 N	77 W	+50	1	2	4	0·15	0·26	0·38	9·0	11·2	10·8
San Francisco	37 N	122 W	+44	1	2	3	0·23	0·38	0·53	8·6	8·7	8·2
Baton Rouge	30 N	91 W	+41	0	2	3	0·26	0·28	0·45	9·0	10·3	8·2
San Juan	18 N	66 W	+30	2	4	4	0·20	0·28	0·34	11·6	11·3	9·6
Maui	21 N	156 W	+21	1	5	6	0·26	0·40	0·53	11·7	11·8	11·1
Huancayo	12 S	75 W	−1	5	5	4	0·18	0·20	0·18	11·3	12·8	13·4
Brisbane	28 S	153 E	−36	4	1	1	0·32	0·28	0·18	9·6	11·2	11·3
Watheroo	30 S	116 E	−42	5	2	1	0·48	0·34	0·23	9·3	10·7	10·7
Canberra	35 S	149 E	−44	4	2	2	0·38	0·23	0·11	8·9	11·3	10·7
Trinidad	11 N	61 W	+22	1		5	0·23		0·38	11·0		10·0
Delhi	29 N	77 E	+19	0		5	0·15		0·45	12·9		10·0
Chungking	29 N	107 E	+18	0		5	0·26		0·48	12·3		11·0
Okinawa	26 N	128 E	+15	1		5	0·18		0·38	11·8		11·2
Calcutta	23 N	89 E	+12	1		6	0·18		0·23	11·8		11·2
Palmyra	6 N	162 W	+6	3		6	0·26		0·32	11·0		12·3
Guam	14 N	145 E	+4	1		6	0·26		0·26	11·8		12·0
Leyte	11 N	125 E	0	3		5	0·23		0·23	11·0		11·4
Johannesburg	26 S	28 E	−27	6		1	0·53		0·32	7·4		10·0

The day time maximum of $h'F2$ is somewhat earlier at sp-min than sp-max, but this is almost certainly due to the effect of group retardation on the sp-min curves. The f^0F1 curves are not shown on the diagrams but the noon values are given and one may easily visualise when and where the f^0F1 and f^0F2 curves will come close to one another. These are the occasions when $h'F2$ is increased by group retardation. They tend to occur before noon.

The only $h'F2$ features that vary conspicuously with the sp-cycle are the sharp maxima near 4 h, 12 h, and 20 h at tropical stations. Apparently the tendency is for a maximum at only one of these times. Similar maxima have been noticed at 20 h in the sp-max curves of Palmyra, Guam, and Singapore, and one might conclude that they are a feature of all localities with a magnetic latitude of less than 10°. The 4 h and 20 h maxima occur more at sp-max than sp-min and do not appear to correlate with any f^0F2 anomalies.

Since the sp-min anomaly has f^0F2 maxima and minima near 4 h and 20 h one might expect at these times to find some sp-min $h'F2$ features that would correlate with D. No definite features have been found, but there is a correlation in the sense that the midnight maximum of $h'F2$ comes later at sp-min than at sp-max when D is positive. This is best demonstrated from our data by comparing $h'F2$ in the evening (18 h–24 h) with the morning (0 h–6 h) at sp-max and sp-min. Measurements are given in Table 5. In Fig. 10 we use the height change index sp-min evening + sp-max morning − sp-min morning − sp-max evening, and

compare it with the sp-min anomaly D. Huancayo and Fairbanks are evidently disturbed by other features but otherwise there is a fairly close relation.

The most consistent feature of the h' curves is the increased mid-day height in summer as compared with winter. Mean curves for mid latitude stations are shown in Fig. 11. There can be little doubt that great summer heights are physically associated with the smallness of f^0F2 at the same time (see Fig. 9). The noon

Table 5. $h'F2$ in the early morning (0h–6h mean) and evening (18h–24h mean) morn. = morning, ev. = evening, diff. = morn-ev, ch. = height change index = sp-min-ev + sp-max-morn − sp-min-morn − sp-max-ev.

Station	sp-cycle	N. Solstice				Equinox				S. Solstice			
		morn	ev	diff	ch	morn	ev	diff	ch	morn	ev	diff	ch
		km											
Fairbanks	sp-max	395	330	+65	+23	370	288	+82	+47	338	272	+65	+13
	sp-min	320	278	+42		318	282	+35		335	282	+52	
Washington	sp-max	272	268	+5	−3	258	250	+8	−10	258	245	+12	−3
	sp-min	272	265	+8		275	258	+18		272	258	+15	
San Francisco	sp-max	288	278	+10	−2	295	255	+40	+30	280	242	+38	+33
	sp-min	265	253	+12		265	255	+10		255	250	+5	
Baton Rouge	sp-max	310	292	+18	−7	290	275	+15	+7	292	265	+28	+35
	sp-min	302	278	+25		290	282	+8		275	282	−7	
Maui	sp-max	250	265	−15	−23	258	245	+13	+13	288	245	+42	+32
	sp-min	258	265	+8		252	252	0		252	242	+10	
Huancayo	sp-max	262	305	−43	−50	255	332	−78	−62	262	342	−80	−58
	sp-min	258	250	+8		252	267	−15		268	290	−22	
Brisbane	sp-max	268	250	+18	+22	272	265	+8	+15	275	288	−12	−7
	sp-min	268	272	−5		262	270	−8		278	282	−5	
Watheroo	sp-max	268	248	+20	+25	275	260	+15	+12	275	268	+8	+8
	sp-min	232	238	−5		245	242	+3		252	252	0	
Canberra	sp-max	265	250	+15	+15	270	260	+10	+8	270	270	0	0
	sp-min	258	258	0		268	265	+2		275	275	0	

values of $h'F2$ correlate negatively with the critical frequency range V, but this is probably only another manifestation of the correlation between $-f^0$ and h_{max}, where h_{max} is the true height of maximum electron density.

Besides the daytime maximum in $h'F2$ Fig. 11 gives indications of a night maximum. This has been studied by MARTYN [4] who found the maximum to occur

VOL. 4 1953

later in the winter than the summer. The time of the maximum cannot be judged accurately from Figs. 1 and 2, but instead the morning (0 h—6 h mean) and evening (18 h–24 h mean) heights have been measured. These are given in Table 5. A positive difference morning–evening, implies that the maximum occurred

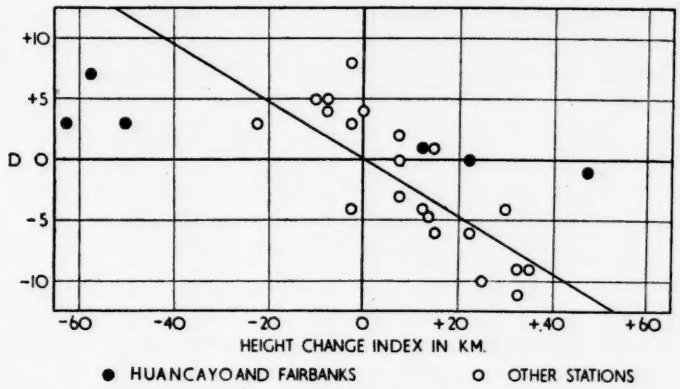


Fig. 10. Change of h' in evening (18 h–24 h) and morning (0 h–6 h) at sp-min as compared with sp-max plotted against sp-min anomaly. Positive height change means sp-min ev + sp-max morn > sp-min morn + sp-max ev.

later than midnight. It is found that at sp-max the winter values are consistently more positive than the summer values (mean excess = 22 km), but at sp-min this effect is zero.

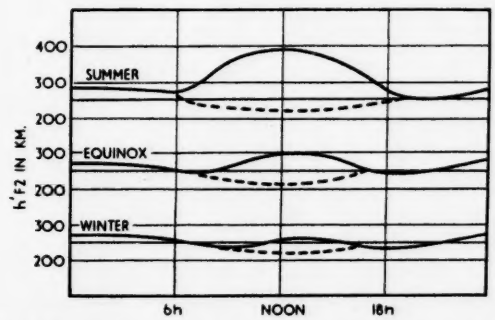


Fig. 11. Diurnal variation of $h'F2$ at mid latitude stations.

The sunrise anomaly S in f^0F2 does not appear to correlate with any h' feature near sunrise. However there is a correlation in the sense that when S is large there is only a small daytime increase of h'.

6. Discussion

There is no doubt that the quantitative measurement of the three anomaly indices D, S, and V, and other world factors of $F2$ displayed in the foregoing sections, could be put to good use if they were understood physically. The observations

do not lend themselves to straightforward explanation, but we will derive what conclusions we can from them.

The sp-min anomaly shows the existence of a world-wide $F2$ layer oscillation that is much greater at sp-min than sp-max. The anomaly is characterised by adjacent decreases and increases of f^0F2 (e.g. sp-min winter evenings). Such an effect cannot be explained by either (a) a sudden supply of electrons, since that would not account for their disappearance just before (or just after) the increase, nor (b) a sudden change in rate of disappearance of electrons, since that would not account for the electron increase just after (or just before) the decrease. It appears rather that the effect must be due to a dilution and concentration of the electrons existing in the layer. Such changes could be made by movements in the ionosphere. If the movements were the same at sp-max as at sp-min then the features of the effect as seen in Figs. 1 and 2, where f^0F2 is on a logarithmic scale, would not change with the sp-cycle. The fact that the features are much stronger at sp-min than sp-max must mean that the movement is much greater at sp-min. The regularity of the features suggests that the movement is an oscillation stimulated by periodic heating or by tidal forces. These would be the same (in any case not less) at sp-max as at sp-min, and the fact that the anomaly is small at sp-max shows there must be a high resistance to the motion at that time. Probably the oscillation is electrically or magnetically damped by the higher charge density or conductivity of sp-max. As evidence supporting this suggestion we note that when the features of the sp-min anomaly are visible in the sp-max curves (e.g. San Francisco winter evening) they are retarded by an hour or more. The anomaly has a reversal of phase between summer and winter and between northern and southern hemispheres which is similar to the 8-hourly ground oscillation of the atmosphere [5]. However in the sp-min anomaly the main maximum and minimum are 8 hours apart and the oscillation does not appear to be a harmonic. It appears rather that the curves of Fig. 5 are composed of a day part having no special tendency to vary but fitted on to a night part that tends to increase or decrease steadily. When night f^0 is decreasing (more rapidly than normal), i.e. when D is positive, h' is increasing and has the effect of making the midnight maximum appear later (relative to $D = 0$ or to sp-max conditions).

The sunrise anomaly appears to have the character of an oscillation or wave associated with sunrise. It does not vary systematically in phase or magnitude with the sunspot cycle and therefore has a different physical mechanism from the sp-min anomaly. One would expect it to be associated with the heating of the ionosphere by solar radiation. At the effective wavelengths the intensity of radiation is greater at sp-max than at sp-min, and this may counterbalance any tendency towards increased electrical damping at sp-max.

There does not appear to be any straightforward way of fitting the observed $F2$ anomalies into MARTYN'S tidal theory [4]. The sp-min anomaly shows some superficial similarity to tidal predictions but there is no agreement in detail and the complete reversal of the anomaly with season is not explained. It is difficult to reconstruct tidal movements that could produce the results found.

The daily variation anomaly V should be explained in terms of the relation between critical frequency and layer height whereby abnormally low f^0F2 is

VOL. 4 195

associated with greater h_{max}. The explanation of this effect constitutes one of the urgent ionospheric problems. There are two extreme outlooks. The first suggests that the $F2$ region is not a proper layer whose lower boundary is caused by exhaustion of ionising radiation, but is the higher region of the $F1$ layer and is distinguished from it by a much lower electron loss rate [6], [7]. The rate of ionisation would decrease with height (being proportional to pressure) and this might account immediately for the decrease of f^0 with h_{max}. The explanation would imply that within the $F2$ layer the electron loss rate is almost independent of height. On the other hand RATCLIFFE's observations [8] would suggest that this view is not necessary. He finds the decrease of f^0 with increased h_{max} can be accounted for by the observation that the layer width increases very greatly with increased h_{max}, and hence, for a given total electron content, f^0 will decrease with increased h_{max} or h'.

We may obtain some information on the rate of electron loss from the fact that the sp-cycle variations of f^0F2 are greater than f^0F1. If we assume with BRADBURY [6] and BATES [7] that both layers are ionised by the same band of radiation, the sp-cycle changes in the ionising radiation must be the same, and the difference in the f^0 change must be due to a difference in the electron loss formula. The results point strongly [1] to an $F2$ electron disappearance rate of $dN_e/dt = -\beta N_e$. This suggests an attachment process, but in that case the disappearance rate would be proportional to gas pressure in disagreement with the f^0–h_{max} relation. The observations therefore point to an electron decay formula that is not supported by recombination theories [9]. A somewhat similar conclusion may be reached by following RATCLIFFE's approach to the $F2$ problem [8].

From the discussion it can be seen that ionospheric observations are not yet amenable to consistent explanation. The results of the present paper do not appear to improve the situation, but they are too regular and world-wide to be neglected.

Acknowledgments—This work was commenced at the Commonwealth Observatory, Canberra, and concluded at the University of London Observatory, Mill Hill. The author wishes to acknowledge his indebtedness to Miss JOYCE DALY for assistance with the reductions, and to Dr. D. F. MARTYN and Mr. J. A. RATCLIFFE for useful discussion.

REFERENCES

[1] ALLEN, C. W.; Terr. Mag. 1948 **53** 433–448. [2] CHAPMAN, S. and BARTELS, J.; "Geomagnetism," p. 505, Oxford, 1940. [3] APPLETON, E.; J. Atm. and Terr. Physics 1951 **1** 106–113. [4] MARTYN, D. F.; Proc. Roy. Soc. A. 1947 **189** 241–260. [5] MITRA, S. K.; "The Upper Atmosphere," 2nd ed., p. 35, Calcutta, 1952. [6] BRADBURY, N. E.; Terr. Mag. 1938 **43** 55. [7] BATES, D. R.; Proc. Roy. Soc., A. 1948 **196** 562. [8] RATCLIFFE, F. A.; J. Geoph. Res. 1951 **56** 487–507. [9] MASSEY, H. S. W.; Phil. Mag. Supp. 1952 **1** 395–426.

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 68 to 77. Pergamon Press Ltd., London

Über die Bedeutung einer weichen Röntgenstrahlung der Sonnenkorona für die Bildung der normalen ionosphärischen *E*-Schicht

The effect of soft X-rays from the solar corona on the formation of the normal ionospheric *E*-layer

GERHARD ELWERT

Astronomisches Institut, Universität Tübingen

(*Received* 25 *April*, 1953)

ABSTRACT

This continues a paper in which the author (1953) calculated the radiation of the solar corona in the spectral region of soft X-rays, as reported in Part I. In Part II, the effect of that radiation in the earth's atmosphere is studied, in particular the height at which the radiation is absorbed, and the number of electrons in case of equilibrium between ionisation and recombination. The results obtained agree well with the normal *E*-layer.

Die vorliegende Untersuchung schließt an eine andere Arbeit des Verfassers (ELWERT, 1953) an, in welcher die Strahlung der Sonnenkorona im Gebiet weicher Röntgenstrahlen berechnet wurde. Hierüber wird zunachst in Teil I berichtet.

Aufgabe der Arbeit ist es, die Wirkung dieser Strahlung in der Erdatmosphäre zu untersuchen. Dabei interessiert besonders die Höhe, in der die Strahlung absorbiert wird und die Zahl der im Gleichgewicht von Ionisation und Rekombination vorhandenen Elektronen. Die erhaltenen Ergebnisse passen gut zur normalen ionosphärischen *E*-Schicht. Dies wird in Teil II ausgeführt.

VOL.
4
1953

1. DIE STRAHLUNG DER SONNENKORONA IM GEBIET WEICHER RÖNTGENSTRAHLEN

Die Berechnung der Strahlung, die der vorliegenden Arbeit zu Grunde gelegt wird, geht aus von unseren Kenntnissen über den Aufbau der Sonnenkorona (SIEDENTOPF, 1950). Diese beziehen sich insbesondere auf die Elektronendichte der Korona, die aus dem Helligkeitsverhältnis Korona-Sonne ermittelt werden konnte sowie ihre Temperatur, wie sie aus dem Intensitätsverhältnis zweier Emissionslinien der Korona zu erschließen ist. Ferner kann für die chemische Zusammensetzung auf Grund der Ergebnisse der quantitativen Spektralanalyse die kosmische Häufigkeit der Elemente verwendet werden.

Für die Theorie der Temperaturbestimmung und der Strahlungsemission sind die Wirkungsquerschnitte der atomphysikalischen Elementarprozesse der Ionisation, Rekombination und Anregung wesentlich. Die ersten beiden gehen bereits in die Ionisationsformel der Sonnenkorona ein. Da die Sahagleichung auf die Korona nicht anwendbar ist, muß nämlich der Mechanismus des Ionisationsgleichgewichts untersucht werden. Im stationären Zustand ist, wovon schon BIERMANN (1947), v. d. R. WOOLLEY und ALLEN (1948) und MIYAMOTO (1950) ausgegangen sind, die Zahl der Stoßionisationen gleich der Zahl der Photorekombinationen. In einer früheren Arbeit des Verfassers (ELWERT, 1952) wurden die Wirkungsquerschnitte für diese Prozesse genauer diskutiert. Bei den Stoßionisationen wurden wellenmechanisch gerechnete und experimentell geprüfte Werte

herangezogen, bei den Photorekombinationen wurden die Beiträge höherer Schalen berücksichtigt. Mit Hilfe der Ionisationsformel wurde dann das Häufigkeitsverhältnis von Fe X zu Fe XIV in Abhängigkeit von der Temperatur ermittelt. Umgekehrt kann aus dem beobachteten Intensitätsverhältnis der roten und grünen Koronalinie auf die Häufigkeit dieser beiden Ionen und damit auf die Temperatur der Korona geschlossen werden. Bei gleicher Intensität der genannten Linien ergibt sich eine Temperatur von etwa $7 \cdot 10^5$ Grad, die je nach dem Intensitätsverhältnis um etwa 10^5 Grad schwankt. Eine Änderung der Temperatur ungefähr um diese Größe findet im 11-jährigen Zyklus der Sonnenflecken statt. Die Temperaturen sind in guter Übereinstimmung mit den aus Messungen der Koronastrahlung im Radiofrequenzgebiet erschlossenen (REULE, 1952). Für $6 \cdot 10^5$ Grad und $7 \cdot 10^5$ Grad wurden dann mit Hilfe der Ionisationsformel Relativwerte der Besetzung der Ionisationszustände von H, He, Ne, O, N, C, Fe, Mg, Si und S ermittelt. Zur Bestimmung der Ionendichte wurde die Häufigkeit dieser Elemente nach der quantitativen Analyse der Sonnenphotosphäre von UNSÖLD (1948) und die Elektronendichte nach der BAUMBACHschen Formel (1931) verwendet. Da die Zahl der Elementarprozesse der Ausstrahlung proportional zur Ionendichte und Elektronendichte ist, sind diese Größen der Ausgangspunkt für die Berechnung der Emission der Korona im Gebiet weicher Röntgenstrahlen; sie wurde in einer anderen Arbeit des Verfassers (ELWERT, 1953) ausführlich untersucht. Infolge der räumlichen Dichteschwankungen in der Korona ist mit einer Unsicherheit um einen Faktor Q zu rechnen, der Werte zwischen 1 und etwa 2 annehmen kann.

Die bei den Photorekombinationen der freien Elektronen mit den hochionisierten Atomen der Korona entstehende kontinuierliche Ausstrahlung ergibt ein Emissionsspektrum, in dem die Kante von He III bei 228 Å und Kanten der schwereren Elemente bei etwa 50 Å hervortreten. Die letzteren überwiegen die Strahlung des Wasserstoffs bei diesen kurzen Wellenlängen um zwei Größenordnungen. Für die Strahlungsintensität des He-Grenzkontinuums wurde in Erdentfernung etwa $8 \cdot 10^{-3}\ Q$ erg/cm² sec, für die der schwereren Elemente etwa die Hälfte gefunden.

Größere Intensität als das Kontinuum enthält die Linienemission. Ihre Anregung kann durch Rekombination in höhere Schalen und durch Elektronenstöße erfolgen, wobei der letztere Vorgang der wirksamere ist. Zur Berechnung der Wirkungsquerschnitte für Stoßanregung wurde von der wellenmechanischen BORNschen Stoßtheorie ausgegangen. Ihre Resultate sind für Stoßionisationen bis zu Elektronenenergien wenig oberhalb der Ionisationsenergie in guter Übereinstimmung mit experimentell aufgenommenen Ionisierungsfunktionen für hohe Ionisationsenergien. Bei der Stoßanregung kann man sich auf Übergänge zwischen dem Grundniveau und der ersten angeregten Schale beschränken. Die Wellenlängen dieser Übergänge wurden für die hochionisierten Atome der Korona auf Grund ihrer Wasserstoffähnlichkeit aus den Ionisationsenergien berechnet und mit den Werten verglichen, die aus den Energiedifferenzen experimentell erschlossener Terme folgen. Bei der Linienemission der Korona wurde schließlich die Selbstabsorption berücksichtigt, während von der Absorption des Kontinuums abgesehen werden kann.

Die Rechnungen führten zu einem Spektrum der Korona, das in Abbildung

1 für $6 \cdot 10^5$ Grad und in Abbildung 2 für $7 \cdot 10^5$ Grad dargestellt ist. Der Hauptbeitrag der Linien liegt in der Nähe von 90 Å; ihr Energieinhalt ist in beiden Fällen in Erdentfernung etwa 0,06 Q' erg/cm² sec. Dabei enthält der Faktor Q′ die bereits angegebene Unsicherheit Q durch die Dichteschwankungen der Korona, wie die etwa ebenso grosse Unsicherheit der Wirkungsquerschnitte für Stoßanregung. Zum Vergleich ist der Beitrag der Grenzkontinua bei den Wellenlängen eingetragen, die den Kantenfrequenzen entsprechen.

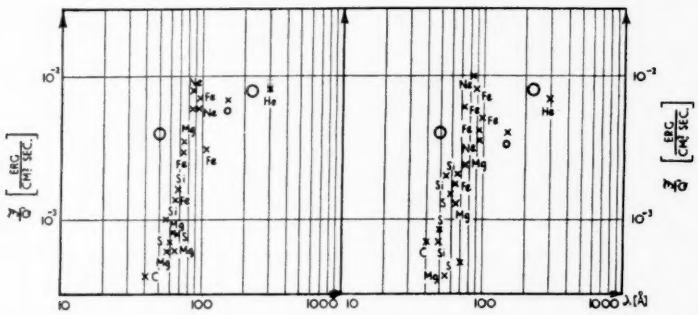


Abb. 1 und 2. Spektrum der Sonnenkorona für $6 \cdot 10^5$ Grad und $7 \cdot 10^5$ Grad. Strahlungsintensität in Erdentfernung I, dividiert durch Q'. Die kontinuierliche Strahlungsintensität ist für die 50 Å—Gruppe und He durch Kreise eingezeichnet.

Fig. 1 and 2. Spectrum of the solar corona for $6 \cdot 10^5$ °K and $7 \cdot 10^5$ °K. Intensity of radiation I in the distance of the earth, divided by Q'. The intensity of the continuum for the 50 Å group and He is denoted by circles.

VOL. 4 1953

2. Bildung der normalen ionosphärischen E-schicht

Es ist heute auf Grund des jahres- und tageszeitlichen Verlaufs der Elektronenkonzentration, ihrer Längen- und Breitenabhängigkeit, sowie auf Grund von ionosphärischen Beobachtungen bei Sonnenfinsternissen als gesichert anzusehen, daß jedenfalls die normale E- und die F_1-Schicht von einer solaren Wellenstrahlung herrühren.* Von dieser sind folgende Eigenschaften zu fordern: sie muß einmal die nötige Energie zur Bildung der Ionen liefern, zum andern muß die Strahlung in der richtigen Höhe wirksam werden; schließlich soll sie die Variation der Elektronendichte im 11-jährigen Zyklus der Sonnenflecken erklären.

Zunächst ist die Wellenlänge der ionisierenden Strahlung durch die Ionisationsenergien der in Betracht zu ziehenden Prozesse nach oben hin begrenzt. Es gilt

$$\begin{aligned}
O_2 &\rightarrow O_2^+ + e \quad \chi \geqq 12{,}2\ \text{eV} \quad \text{d.h. } \lambda \leqq 1000\ \text{Å} \\
O &\rightarrow O^+ + e \quad \geqq 13{,}6\ \text{eV} \quad \leqq 910\ \text{Å} \\
N_2 &\rightarrow N_2^+ + e \quad \geqq 15{,}5\ \text{eV} \quad \leqq 795\ \text{Å} \\
N &\rightarrow N^+ + e \quad \geqq 14{,}5\ \text{eV} \quad \leqq 850\ \text{Å}
\end{aligned}$$

Es kommen also von vornherein nur Wellenlängen unterhalb von 1000 Å in Betracht. Die von der Photosphäre gelieferte Strahlungsintensität im fernen Ultraviolett unterhalb dieser Wellenlänge ist nun, wie schon seit längerer Zeit

* (vgl. Zenneck, 1949, Rawer, 1953)

bekannt ist, um mehrere Zehnerpotenzen kleiner, als zur Bildung der ionosphärischen Schichten erforderlich wäre. Deshalb hat KIEPENHEUER (1945) die Strahlung der Sonnenkorona in diesem Wellenlängenbereich untersucht, dabei aber nur die kontinuierliche Emission des Wasserstoffs in Betracht gezogen. Er erhielt daher selbst bei einer angenommenen Vergrößerung der Emission um zwei Zehnerpotenzen, die einer Erhöhung der BAUMBACHschen Elektronendichte um den Faktor 10 entspricht, nicht die erforderliche Strahlungsenergie. V. D. R. WOOLLEY und ALLEN (1948) fanden dagegen bei einer Abschätzung der Linienemission einiger Elementgruppen mit einem Ansatz für den Wirkungsquerschnitt für Stoßanregung eine sehr viel höhere Gesamtemission der Korona unterhalb 1000 Å von $6 \cdot 10^{26}$ erg/sec, die einer Intensität in Erdentfernung von 0,2 erg/cm² sec entspricht. Der Hauptbeitrag kommt dabei von einem Übergang innerhalb der L-Schale von MgX, für den die beiden Autoren auf Grund einer Extrapolation der Termdifferenzen Lithium-ähnlicher Spektren eine Energiedifferenz von 20 eV fanden, die einer Wellenlänge von etwa 600 Å gleichkommt. Diese Strahlung würde also in dem Energiebereich wenig unterhalb der obengenannten Ionisationsenergien liegen und damit etwa den Vorstellungen über die wirksamen Wellenlängen von BHAR (1938) entsprechen. Die von V. D. R. WOOLLEY und ALLEN berechnete Strahlungsenergie würde andererseits die erforderliche erheblich übertreffen. Dies liegt offenbar daran, daß der von diesen Autoren verwendete Ansatz für den Wirkungsquerschnitt zu hohe Werte liefert, wie auch im Fall des Übergangs zwischen verschiedenen Schalen durch wellenmechanische Rechnungen (ELWERT, 1953) gefunden wurde. Wenn sich eine Strahlung von etwa 600 Å möglicherweise bei genauerer Rechnung in der erforderlichen Intensität ergibt, so würde sie aber zum Verständnis der *E*-Schicht nicht geeignet sein, da sie schon in größerer Höhe absorbiert wird und damit der zweiten der oben gestellten Bedingungen nicht genügt.

Ionosphärische Beobachtungen ergeben für die E-Schicht eine Höhe von etwa 120 km. Kennt man die durchsetzte Luftmasse und den von der Wellenlänge abhängigen Absorptionskoeffizienten für Luft, so können Rückschlüsse auf die Frequenz der ionisierenden Strahlung gezogen werden. Unter der Annahme einer Partikeldichte in der *E*-Schicht von $6 \cdot 10^{12}$/cm³, wie sie aus Modellvorstellungen der hohen Atmosphäre folgte, wurde von SIEDENTOPF (1948, 1950) und von HOYLE und BATES (1948) auf eine Strahlung von etwa 30 Å geschlossen. Neuerdings konnten nun bei Aufstiegen von V2-Raketen von HAVENS, KOLL und LA GOW (1952) experimentelle Werte für den Druck in der hohen Atmosphäre bis zu 160 km Höhe gewonnen werden, aus denen auch Dichte und Temperatur folgen. Diese Aufstiege fanden in White Sands in New Mexiko in 32° nördlicher Breite und 106° westlicher Länge statt. Die erhaltenen Resultate sind in der folgenden Tabelle für die interessierenden Höhen angegeben.

In der Höhe der *E*-Schicht ist die Dichte also merklich geringer, als erwartet wurde. Infolgedessen würde eine Strahlung von 30 Å zu tief eindringen; die Ionisationsursache muß also eine Strahlung etwas größerer Wellenlänge sein. Sie soll nun unter Verwendung der oben angegebenen Drucke ermittelt werden.

Die in einer bestimmten Höhe h in einer Schicht der Dicke Δh absorbierte Energie ΔE ist proportional zur Strahlungsintensität I in dieser Höhe, zum

Höhe Km	*Druck mmHg*	*Dichte g/m³*	*Moleküldichte N cm⁻³*	*Temperatur °K* (N_2O_2)	(N_2,O)	*äquiv. Schichtdicke cm*
80	$1{,}0\cdot10^{-2}$	$2{,}5\cdot10^{-2}$	$6\cdot10^{14}$	190	...	10
90	$1{,}9\cdot10^{-3}$	$4{,}0\cdot10^{-3}$	10^{14}	210	...	2
100	$4{,}2\cdot10^{-4}$	$8{,}0\cdot10^{-4}$	$2\cdot10^{13}$	240	...	0,44
110	$1{,}2\cdot10^{-4}$	$2{,}0\cdot10^{-4}$	$4\cdot10^{12}$	270	220	0,13
120	$3{,}5\cdot10^{-5}$	$5{,}0\cdot10^{-5}$	10^{12}	330	270	$3{,}7\cdot10^{-2}$
130	$1{,}5\cdot10^{-5}$	$2{,}0\cdot10^{-5}$	$5\cdot10^{11}$	390	320	$1{,}6\cdot10^{-2}$
140	$7{,}0\cdot10^{-6}$	$7{,}0\cdot10^{-6}$	$2\cdot10^{11}$	450	370	$7{,}4\cdot10^{-3}$
150	$3{,}0\cdot10^{-6}$	$3{,}0\cdot10^{-6}$	$7\cdot10^{10}$	510	420	$3{,}2\cdot10^{-3}$

Massenabsorptionskoeffizienten K und der Dichte ρ. Führt man den Absorptionskoeffizienten je cm durchstrahlte Schichtdicke von Luft unter Normalbedingungen κ ein, so kann man auch schreiben

$$\Delta E = K\rho I\Delta h = \kappa(\rho/\rho_0)I\Delta h \qquad (1)$$

Im Falle einer Atmosphäre konstanter Temperatur erhält man eine CHAPMAN-Schicht; bei ihr ergibt sich maximale Absorption in der Höhe, in der die Strahlungsintensität auf den e-ten Teil ihres extraterrestrischen Wertes abgesunken ist (CHAPMAN, 1931). Da die Temperatur innerhalb einer Schicht von etwa 10 km Dicke, in der sich das Maximum ausbildet, nicht sehr stark variiert, gilt dies in guter Näherung auch für eine nicht isotherme Atmosphäre. Die Höhe des Absorptionsmaximums ist also durch die Bedingung bestimmt, daß das Produkt aus durchstrahlter Materie und Absorptionskoeffizient 1 ergibt.

VOL. 4 1953

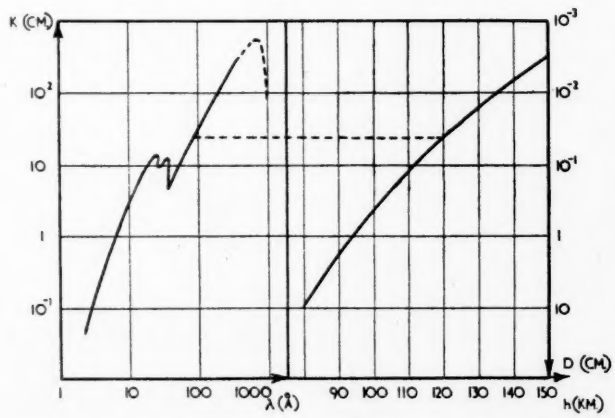


Abb. 3. Absorptionskoeffizient κ, bezogen auf 1 cm Luft unter Normalbedingungen und effektive Schichtdicke D der Atmosphäre. Aus der Abb. ist der Zusammenhang zwischen Wellenlänge und Höhe maximaler Absorption zu ersehen.

Fig. 3. Coefficient of absorption κ for a layer of 1 cm of air under normal conditions and effective thickness of the atmosphere. From the figure the connection between wavelength and height of greatest absorption can be seen.

Nun ist die durchstrahlte Luftmasse durch den Druck und den Einfallswinkel gegeben. Rechnet man den Druck auf die Dicke einer Luftsäule unter Normalbedingungen D um, so erhält man die in der letzten Spalte der Tabelle aufgeführten Werte. Sie ergeben die auf der rechten Seite von Abb. 3 eingezeichnete Kurve. Andererseits ist der experimentelle Absorptionskoeffizient für Luft nach SCHNEIDER (1940) und SIEDENTOPF (1948, 1950) in Abhängigkeit von der Wellenlänge auf der linken Seite derselben Abbildung wiedergegeben. Er ist bezogen auf Luft unter Normalbedingungen von 1 cm Schichtdicke. Für einen Sonnenstand von 45° ist er noch um einen Faktor $\sqrt{2}$ gegenüber der SIEDENTOPF'schen Kurve vergrößert. Sowohl die Luftmasse als der Absorptionskoeffizient sind in demselben logarithmischen Mass aufgetragen und zwar letzterer nach oben positiv, erstere nach unten positiv. Es sind also für jede Ordinate die Logarithmen der beiden Größen entgegengesetzt gleich. Geht man nun von einer bestimmten Wellenlänge aus, so kann man zunächst links den Absorptionskoeffizienten ablesen. Sein reziproker Wert erscheint auf der rechten Ordinate und ergibt zugleich die Schichtdicke, die die Strahlung durchsetzen muß, bis maximale Absorption auftritt. Die dazugehörige Höhe der Atmosphäre kann auf der rechten Abszisse abgelesen werden. Wie man sofort sieht, wird Strahlung einer Wellenlänge von 600 Å weit oberhalb der E-Schicht absorbiert; umgekehrt führt eine Höhe von 120 km auf eine Wellenlänge von etwa 90 Å. Die bei der Berechnung der Koronastrahlung gefundene Wellenlänge maximaler Intensität wird also gerade in der Höhe der E-Schicht Ionen bilden.

Die Absorption in Abhängigkeit von der Höhe kann nach Formel (1) bestimmt werden, und zwar empfiehlt sich die Anwendung eines einfachen graphischen Verfahrens, da die CHAPMANschen Absorptionskurven nur für eine isotherme Atmosphäre gelten. Aus der effektiven Schichtdicke oberhalb einer bestimmten Höhe D und dem bekannten Absorptionskoeffizienten ist zunächst die Strahlungsintensität

$$I = I_0 e^{-\kappa D} \tag{2}$$

zu berechnen. Sie ist unter Verwendung des Absorptionskoeffizienten $\kappa \approx 25$ für Strahlung von etwa 90 Å in logarithmischem Maß in Abhängigkeit von der Höhe in Abb. 4 dargestellt. Außerdem ist in derselben Abbildung ebenfalls in logarithmischem Maß die Dichte der Luft in Einheiten ihrer Dichte am Erdboden ρ/ρ_0 eingetragen. Die in einer bestimmten Höhe absorbierte Energie ist proportional zu beiden Größen und ergibt sich somit bis auf einen konstanten Faktor durch Addition der beiden genannten Kurven. Da in großen Höhen die Materiedichte und in kleinen die Strahlungsdichte verschwindet, entsteht eine Kurve Σ_1 mit einem Maximum, das in etwa 120 km Höhe liegt; dort ist die Intensität, wie aus der Abbildung ebenfalls zu ersehen ist, auf $1/e$ ihres extraterrestrischen Wertes abgesunken. Diese Kurve Σ_1 gilt für das Wellenlängenintervall zwischen 80 und 100 Å, das den Hauptteil der Strahlungsenergie enthält, nämlich bei $7 \cdot 10^5$ Grad etwa $3 \cdot 10^{-2}\, Q'$ erg/cm² sec. In entsprechender Weise können die Kurven für zwei weitere Wellenlängengruppen ermittelt werden und zwar für den Wellenlängenbereich zwischen 80 und 60 Å mit einem Energieinhalt von etwa $1 \cdot 10^{-2}\, Q'$ erg/cm² sec und den Bereich zwischen 40 und 60 Å, der zusammen mit der Strahlung des

Kontinuums etwa dieselbe Energie enthält. Die Summenkurven Σ_2 und Σ_3 für diese Wellenlängengebiete sind in Abbildung 4 ebenfalls eingetragen. Überträgt man die 3 Kurven auf linearen Maßstab und geht durch Multiplikation mit dem Absorptionskoeffizienten κ zur absorbierten Energie über, so erhält man die in Abb. 5 wiedergegebenen Absorptionskurven. Ihre Addition ergibt schließlich

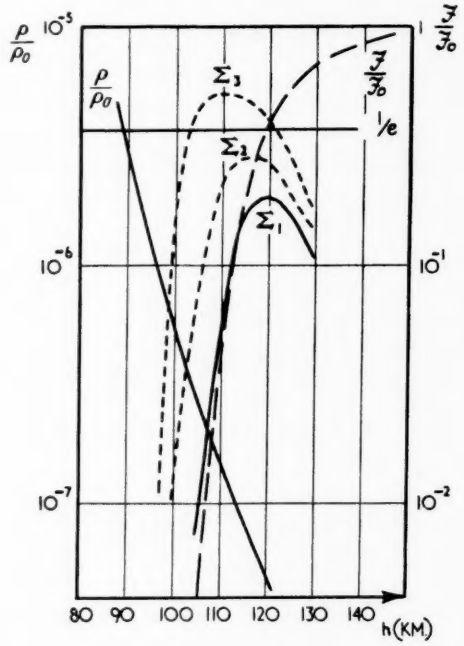


Abb. 4. Absorptionskurven.
Relative Dichte der Luft ρ/ρ_0 ——
Relative Strahlungsintensität I/I_0 für 90 Å — — — —
Summenkurve Σ_1 für 80–100 Å, Σ_2 für 60–80 Å, Σ_3 für 40–60 Å

Fig. 4. Curves of absorption.
Relative density of the air ρ/ρ_0
Relative intensity of radiation I/I_0 for 90 Å — — — —
Sum total Σ_1 for 80–100 Å, Σ_2 for 60–80 Å, Σ_3 for 40–60 Å

die Gesamtabsorption. Dabei ist noch die langwelligere Strahlung von He und O berücksichtigt, die den Abfall oberhalb von 120 km etwas abflacht. Auf ihn erfolgt natürlich durch die in größerer Höhe absorbierte Strahlung wieder ein Anstieg. Diese längeren Wellen werden hauptsächlich bei Elektronenübergängen innerhalb der Schalen gleicher Hauptquantenzahl und von dem Übergangsgebiet Chromosphäre-Korona emittiert.

Es ist weiter zu beachten, daß die Kurven zunächst nur für die absorbierte Strahlungsenergie gelten. Die Photonen haben jedoch bei einer Wellenlänge ven 90 Å eine Energie von etwa 140 eV und bei der kurzwelligsten Gruppe eine ungefähr doppelt so grosse Energie. Andererseits liegen die Ionisationsenergien um 15 eV. Bei den Photoprozessen werden also Elektronen frei, deren Energie groß gegen die Ionisationsenergie ist, die also ihrerseits noch mehrfache Stoßionisationen hervorrufen können, Nun ist der Wirkungsquerschnitt für Stoßionisation

VOL. 4 1953

eines Atoms von der Größenordnung der Fläche der ersten Bohrschen Wasserstoffbahn $\pi a^2 \approx 10^{-16}$ cm^2. Dies gilt auch in ausreichender Näherung für Luftmoleküle. Die Molekülzahl in 120 km Höhe ist nach der Tabelle etwa 10^{12}, die freie Weglänge der Elektronen somit von der Ordnung 10^4 cm; in geringeren Höhen ist sie wegen der größeren Dichte noch kleiner. Die Absorptionskurve für die gesamte Energie wird sich also von derjenigen für die einfallende Strahlung praktisch nicht unterscheiden. Diese kann direkt als Kurve für die Trägerproduktion betrachtet werden.

Zur Bestimmung der Zahl der gebildeten Elektronen selbst werde nun die im Scheitel in einer Schicht von 1 cm Dicke absorbierte Energie ΔE berechnet.

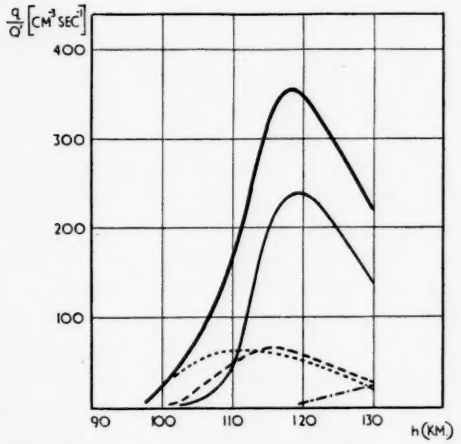


Abb. 5. Elektronenproduktion q in Abhängigkeit von der Höhe, dividiert durch Q'
Strahlung von 80–100 Å ——, 60–80 Å — — — —, 40–60 Å – – – – – Beitrag der längeren Wellen —·—·—·—, Summenkurve ———

Fig. 5. Production of electrons q in function of the height over sea-level, divided by Q'.
Radiation of 80–100 Å ——, 60–80 Å — — — —, 40–60 Å – – – – –, contribution of the longer waves —·—·—·— sum total of the curves ———

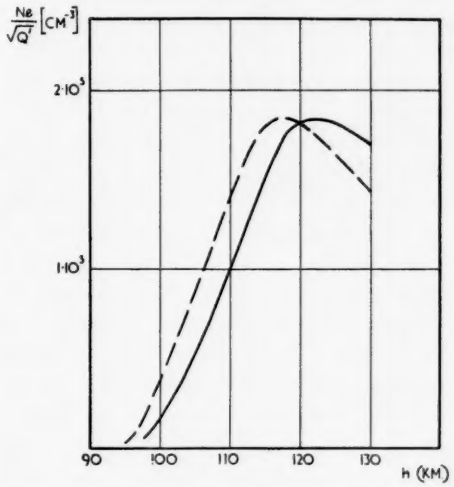


Abb. 6. Elektronendichte N_e in Abhängigkeit von der Höne, dividiert durch $\sqrt{Q'}$. α konstant —— ——, $\alpha \sim \sqrt{N}$ ———

Fig. 6. Density of electrons N_e in function of the height, divided by $\sqrt{Q'}$. α constant—— ——, $\alpha \sim \sqrt{N}$ ———

Dort ist $\rho/\rho_0 \approx 4 \cdot 10^{-8}$, $I = I_0/e$. Mit $\kappa \approx 25$ für Strahlung von etwa 90 Å erhält man $\Delta E \approx 4 \cdot 10^{-7} I$. Beim Übergang zur gebildeten Elektronenzahl ist noch zu berücksichtigen, daß nicht die gesamte Energie der primären Photonen zur Ionisation verwendet wird, da sie zum Teil nur Anregungen und Dissoziationen hervorruft. Nach experimentellen Untersuchungen von KULENKAMPFF (1926) sowie KOSSEL und STEENBECK (1927) und STEENBECK (1928) ist der mittlere Energieverbrauch pro gebildetes Ionenpaar in Luft etwa 30 eV, entsprechend einer Energie von $5 \cdot 10^{-11}$ erg. Die Zahl q_s der im Scheitel je cm^3 und sec gebildeten Elektronen ergibt sich damit zu etwa $350\,Q'$. Andererseits kann die Trägerproduktion aus Echobeobachtungen bestimmt werden. Die Anzahl der direkten Rekombinationen ist proportional zur Elektronendichte und der Ionendichte. Im Gleichgewichtszustand gilt dann bei Vernachlässigung komplizierterer Vorgänge

$$q = \alpha N_e^2 \tag{3}$$

Hierin ist α der Rekombinationskoeffizient in der E-Schicht, für den jedoch keine sehr genauen Angaben bekannt sind. ZENNECK (1949) verwendet den Wert $\alpha = 1 \cdot 10^{-8}$ cm^3 sec^{-1}, während HULBURT (1939) $\alpha = 2 \cdot 10^{-8}$ cm^3 sec^{-1} ermittelt. Die aus den Grenzfrequenzen berechneten Elektronendichten im Scheitel sind nach ZENNECK im Winter etwa $1 \cdot 10^5$ cm^{-3} und steigen im Sommer ungefähr auf $1{,}6\ 10^5$ cm^{-3}. Für die Trägerproduktion q_s erhält er mit dem obengenannten Rekombinationskoeffizienten $q_s = 258$ Ionenpaare/cm^3; diese Zahl würde sich mit dem Hulburtschen α verdoppeln. Andererseits ist zu beachten, daß der Faktor Q' schon allein durch die Dichteschwankungen Werte bis etwa 2 annehmen kann. Die Energie der Röntgenstrahlung der Sonnenkorona ist also innerhalb der Unsicherheiten gerade von der erforderlichen Größe.

Zu den räumlichen Schwankungen der Elektronendichte in der Korona kommen außerdem noch ihre zeitlichen Variationen im 11-jährigen Zyklus der Sonnenflecken hinzu. Nach VAN DE HULST (1947) ist die Elektronendichte im Sonnenfleckenmaximum etwa um den Faktor 1,78 größer als im Minimum. Nun ist die Strahlungsintensität bei konstanter Temperatur der Korona proportional zum Quadrat ihrer Elektronendichte. Zwar variiert die Temperatur im Zyklus etwa um 10^5 Grad, doch wird dadurch praktisch keine Änderung der Energieemission hervorgerufen. Andererseits ist die Zahl der in der E-Schicht gebildeten Elektronen nach (3) proportional zur Wurzel aus der Trägerproduktion q, so daß sich für die Schwankung von N_e wieder etwa der obige Faktor ergibt. Die im 11-jährigen Zyklus beobachtete Variation der Elektronendichte in der E-Schicht ist etwa von dieser Größe. Nach einer Formel von HARNISCHMACHER (1950) ergibt sich ein Faktor 1,4 in 30° geographischer Breite und 1,6 in 50° Breite. Für eine Diskussion der Breitenabhängigkeit wäre eine Kenntnis des Dichteverlaufs in der Erdatmosphäre für verschiedene geographische Breiten erforderlich, die zur Zeit noch nicht vorliegt.

Die in Abb. 5 wiedergegebene Kurve für die Trägerproduktion gibt nun diejenige für die Elektronendichte noch nicht unmittelbar wieder, da zu den direkten Rekombinationen noch kompliziertere Vorgänge, wie Anlagerung an Atome und Rekombinationen in Verbindung mit Molekülddissoziationen hinzukommen. Sie können in der obigen Formel erfaßt werden, wenn α als schwach von der Molekülddichte N abhängig angesehen wird und zwar liegt die Variation des Rekombinationskoeffizienten nach ARGENCE, MAYOT und RAWER (1950) in der Nähe der Proportionalität mit $\sqrt{N}$. Unter Verwendung der Molekülddichten nach HAVENS, KOLL und LA GOW und mit $\alpha = 1 \cdot 10^{-8}$ cm^3 sec^{-1} für 120 km Höhe wurden die Kurven für die Elektronendichte N_e für konstantes α und $\alpha \sim \sqrt{N}$ berechnet und in Abb. 6 dargestellt. Die gültige Kurve dürfte dann zwischen den gezeichneten Kurven liegen. Bei beiden ergibt sich eine halbe Schichtdicke von etwa 20 km, während die Zunahme von α mit wachsendem N eine geringfügige Verschiebung des Maximums nach oben bewirkt, das jedoch in einer Höhe von etwa 120 km bleibt. Die auf Grund der verschiedenen theoretischen und experimentellen Ergebnisse erschlossene Kurve für N_e stimmt also nicht nur bezüglich der Höhe des Maximums, sondern auch in ihrer Breite gut mit den durch Echolotungen ermittelten Werten für diese Größen überein. Auch für die Absolutwerte der Elektronendichte erhält man soweit quantitative

VOL.
4
195

Übereinstimmung, als dies bei der Vielzahl der in die Theorie eingehenden Einzelüberlegungen, dem zugrunde gelegten Koronamodell, der Unsicherheit des Dichteschwankungsfaktors, der Wirkungsquerschnitte für Stoßanregung und des Rekombinations-koeffizienten erwartet werden kann.

Herrn Prof. SIEDENTOPF möchte ich für die Anregung zu der vorliegenden Untersuchung und Diskussionen herzlich danken.

LITERATURVERZEICHNIS

E. ARGENCE, M. MAYOT et K. RAWER	1950	Ann. Géophys. **6**, 242
S. BAUMBACH	1937	Astron. Nachr. **263**, 121
J. N. BHAR	1938	Indian. Journ. of Phys. **12**, 363
L. BIERMANN	1947	Naturwiss. **34**, 87
G. ELWERT	1953	Z. Naturforschung **7a**
G. ELWERT	1952	Z. Naturforschung **7a**, 432
S. CHAPMAN	1931	Proc. Phys. Soc. London **43**, 26
E. HARNISCHMACHER	1950	Compt. Rend. Paris **230**, 1301
R. J. HAVENS, R. T. KOLL and H. E. LA GOW	1952	J. Geophys. Res. **57**, 59
F. HOYLE and D. R. BATES	1948	Terr. Magn. and Atm. Electr. **51**
E. O. HULBURT	1939	Phys. Rev. **55**, 639
H. C. VAN DE HULST	1947	Astrophys. J. **105**, 471
K. O. KIEPENHEUER	1945	Ann. d'Astrophys. **8**, fasc. 3-4
H. KULENKAMPFF	1926	Ann. Phys. **79**, 97
W. KOSSEL und M. STEENBECK	1927	Z. Physik **42**, 832
S. MIYAMOTO	1950	Publ. Astronom. Soc. Japan I
A. REULE	1952	Z. Naturforschung, **7a**, 234
E. G. SCHNEIDER	1940	J. Opt. Soc. Amer. **30**, 128
H. SIEDENTOPF	1950	Die Sonnenkorona, Erg.ex.Nat.-Wiss. XXIII, Springer Berlin-Göttingen-Heidelberg
H. SIEDENTOPF	1950	Grundriß der Astrophysik, Wissenschaftl. Verlagsges. Stuttgart, S. 69
	1948	Naturwiss. **35**, 289
M. STEENBECK	1928	Ann. Phys. **87**, 811
A. UNSÖLD	1948	Z. Astrophysik **24**, 306
R. v. d. R. WOOLLEY and C. ALLEN	1948	Monthly Not. Roy. Astronom. Soc. **108**, 292
J. ZENNECK	1949	Ionosphäre, Erg. ex. Natwiss. XXII, 263
K. RAWER	1953	Die Ionosphäre, P. Noordhoff N. V. Groningen

RESEARCH NOTES

A universal time component in geomagnetic disturbance

(*Received* 9 *June* 1953)

ABSTRACT

Evidence is obtained of a Universal Time component in the diurnal variation of geomagnetic disturbance from the analysis of K-index figures for selected observatories. A possible interpretation of the results is given in terms of the position of the earth's magnetic axis relative to the system of sun and earth and its associated planes.

The diurnal variation of geomagnetic disturbance, as defined by the range of fluctuations of the earth's field, has been shown by various authors to be, over most of the earth and to a close approximation, a *local time* (L.T.) phenomenon with morning minimum and evening maximum. It has been suggested, however, notably by CHAPMAN (1919) and BARTELS (1925) that a component of disturbance may depend on *universal time* (U.T.); and KWEI (1936), using the results of an analysis made by STAGG (1935*a*, *b*), obtained evidence which suggested that such a component does in fact exist. We present here independent evidence of a U.T. component of disturbance.

The existence of a U.T. component in the diurnal variation of magnetic disturbance would imply some degree of rotational asymmetry in the earth. Such asymmetry is readily supplied, for instance, by the obliquity of the earth's magnetic axis, so that the incidence of disturbance may to some extent depend on the position of the earth's magnetic axis relative to the system of sun and earth and its associated planes, viz., the plane of the ecliptic in which the earth moves, and the plane of the sun's equator in which radially emitted solar particles will tend to lie. Any U.T. component may thus depend on one or more of three parameters, namely the angles between the earth's magnetic axis and

(i) the plane of the ecliptic (ϕ_1 say)

(ii) the plane of the sun's equator (ϕ_2)

(iii) the line joining earth and sun (ϕ_3)

(It would be possible to list further factors which might, less plausibly, exert a U.T. influence on magnetic disturbance).

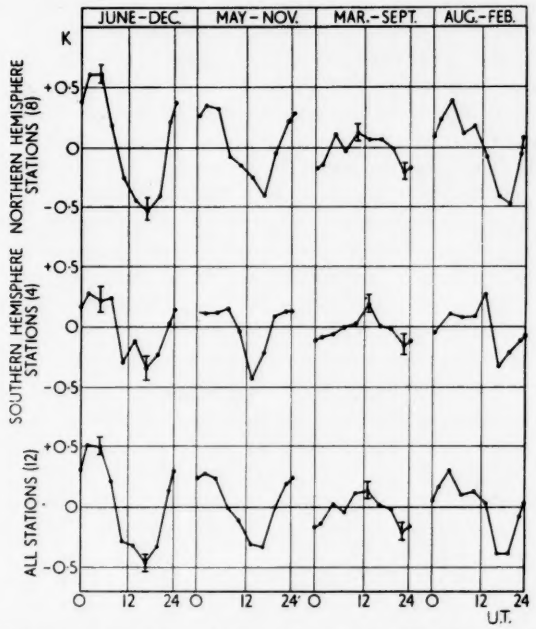
Since the directions of the earth's rotational axis and of the planes of the ecliptic and solar equator are all fixed in space, both ϕ_1 and ϕ_2 vary diurnally between limits which are constant throughout the year. However, with respect to the solar (not sidereal) day, the variations of either angle on two days separated by six months are exactly out of phase. Thus any U.T. variation depending on ϕ_1 or ϕ_2 would be revealed by a significant difference between the diurnal variations of disturbance at any two times of the year six months apart.

The variation of ϕ_3 is more complicated than that of ϕ_1 or ϕ_2, but it can be seen that the variations at the two solstices are diurnal and exactly out of phase, while those at the two equinoxes are in phase. A U.T. variation depending on ϕ_3 should thus give rise to a difference between the diurnal variations of disturbance at the solstices, but not at the equinoxes.

The obvious way to separate any U.T. daily variation of disturbance from the predominant L.T. variation is to choose stations distributed evenly round the earth on a parallel of latitude, and find their mean diurnal variation referred to U.T. Available stations are not sufficiently well distributed for this method to be followed but we can approximate to it by averaging round the world (though not on a circle of latitude). We also adopt the artifice of forming, at each station, differences between daily variations of

VOL. 4 1953

disturbance at intervals of six months, thus removing much the greater part of the L.T. variation. A systematic six-monthly change in the L.T. variation is conceivable, at least between "winter" and "summer," but this would be relatively small and we should hope largely to eliminate it by averaging in U.T. at stations well distributed in longitude. Thus, if the diurnal incidence of disturbance is governed solely by L.T. we should expect such



Mean differences between diurnal variations of magnetic disturbances (in K units) for periods of one month separated by six months (8 northern and 4 southern hemisphere stations): periods are June, December, March, September, and four periods of 31 days centred on 1 May, 1 November, 1 August, 1 February.

average differences to be merely random fluctuations about zero: if, on the other hand, there is a U.T. variation which depends on any (or all) of ϕ_1, ϕ_2, ϕ_3, there should be a significant diurnal variation difference between the two solstices (approx. June-December) of amount equal to twice the average U.T. contribution at the stations considered. We should similarly expect to find a significant average difference between the equinoxes (approx. March-September) if there exists a U.T. component dependent on ϕ_1 or ϕ_2, but no significant March-September difference if any such component either does not exist or depends solely on ϕ_3.

A selection was made of magnetic observatories in both hemispheres with available data of K 3-hour-range disturbance indices, and which were fairly well distributed in longitude such that $\Sigma \sin \Lambda = 0$ in each hemisphere, Λ being geomagnetic longitude. Eight northern and four southern hemisphere observatories were selected, viz. College (+65, 255), Sodankyla (+64, 120) Lerwick (+63, 89), Meanook (+62, 301), Cheltenham (+50, 351), Kakioka (+26, 206), Zo-Se (+20, 189) Alibag (+9, 144); and Pilar (−20, 4), Hermanus (−34, 82), Watheroo (−42, 186), Amberley (−48, 253). The daily variation of K-index was averaged at each station for 60 selected days, mainly in the period 1947–51, for each of the months March, June, September, December; and also for intermediate "months" centred on 1st February, 1st May, 1st August, 1st October. Attention was

confined to days of more-than-average disturbance (daily sum of planetary K (i.e. K_p) indices >20) because of possible inaccuracies in scaling the K-indices at low levels of disturbance. For each station, differences between the daily variations of K-indices were calculated for the four pairs of "months" separated by six months, i.e., June-December, 1 May—1 November, March-September, 1 August–1 February; and these differences were combined in U.T. to give mean differences for the northern and southern hemispheres separately and also for the 12 stations combined. Results are illustrated in the accompanying graph in which the standard errors of selected mean differences, calculated from the standard deviations of the individual K-indices, are shown.

The graphs indicate U.T. disturbance differences which depart significantly from zero in all four pairs of months. Evidence that the variations are not of a seasonal nature is contained in the fact that the solstitial differences are similar in both hemispheres. The difference variation has maximum amplitude for June-December, decreases somewhat in amplitude but retains almost the same phase for May-November and August-February, and has smallest amplitude with changed phase for March-September. This last variation, though small, appeared very probably significant from the original sample of 60 selected days, and the result was tested and verified by using twice as much data for most of the stations: the plotted March-September curves are those for this bigger sample. We have also tested the reality of this result by calculating the mean K daily variation of March-September differences arranged in U.T., at seven further selected northern hemisphere observatories which though not very evenly distributed in longitude, were yet such that $\Sigma \sin \Lambda = 0$. All days of the three available years 1949–51 were used, and the mean variation was found to agree closely with the plotted March-September curves.

The graphs show clearly that the daily variation of disturbance has indeed a U.T. component, the amplitude and phase of which vary throughout the year. It can reasonably be interpreted as consisting of two parts:

(i) A part depending on ϕ_1 or ϕ_2 or both, with no annual variation of amplitude, and time of daily maximum coinciding with maximum ϕ_1, i.e., 0430 U.T. in June and 1030 U.T. in March. (The north magnetic pole is $4\frac{1}{2}$ hours west of Greenwich.)

(ii) A part depending on ϕ_3, having opposite phases in June and December and the same phase in March and September. The time of June and December maximum coincides with maximum ϕ_3, i.e., 0430 U.T. in June, and 1630 U.T. in December.

In the resultant 6-month differences (ii) reinforces (i) in June and December and does not affect it in March and September.

Meteorological Office, Edinburgh.
May 1953. R. P. Waldo Lewis and D. H. McIntosh.

VOL. 4 1953

References

Bartels, J.	1925	Meteorol. Z. **42,** 147
Chapman, S.	1919	Phil. Trans. R. Soc. A. **218,** 1
Kwei, C. T.	1936	Terr. Mag. **41,** 57
Stagg, J. M.	1935*a*	Proc. R. Soc. A. **49,** 298;
	1935*b*	**152,** 277

A note on auroral interaction

(*Received* 23 *May* 1953)

It has been noted on many occasions that the aurora polaris interacts with radio waves directed towards it to cause a modulation of the reflected signals. The amount of the modulation, measured when a probing radio wave of 50 mc/sec was employed, varied between 50–200 c/sec, and the depth of modulation was irregular. When radiotelephone communication was attempted by means of auroral interaction, the reflected signal has been described as having a characteristic flutter; the "voice" is heard against a background which contains rumbling tones, deep hissing, etc. At times severe garbling is encountered, and no intelligence can be extracted from the auroral reflected signal. Because of the reasoning given below, it is believed that the modulation caused by auroral interaction probably arises because of a Doppler effect.

Near 100 km over the temperate latitudes, observations on the movement of sporadic *E* ionization and the movement of ionospheric irregularities, indicate the existence of winds

Table 1. *Radio wave modulation* (*c*/*sec*) *expected from a moving auroral reflecting screen*

Speed of Reflecting Curtain (*cm*/*sec*)	*Radio Probing Frequency* (*mc*/*sec.*)		
	10	50	100
7×10^3	4·7	24	47
3×10^8	2×10^5	10^6	2×10^6

having a velocity of about 250 km/hr. If the auroral reflecting screen moves with this speed along the path of the probing radio wave, a Doppler shift of 5–47 c/sec would be expected. On the other extreme, protons which produce at least some types of aurorae move with velocities of about 3×10^8 cm/sec. Reflection of radio waves by the incoming surface of the incident protons (or by the incoming ionization surface which the protons produce) would cause a Doppler effect of about 10^5–10^6 c/sec. The modulation produced by the auroral reflecting surface moving with the wind is too small by a factor of 5–50 to clarify the observed results; the Doppler shift from the surface of incoming protons is much higher than has yet been measured. The magnitude of the Doppler effect expected with probing radio frequencies of 10, 50 and 100 mc/sec for various speeds of a reflecting curtain moving along the ray path is summarized in Table 1. The results may be compared with those of Table 2 which indicate the speeds of the auroral reflecting curtain necessary to produce a modulation of 50–200 c/sec.

While neither of the above described mechanisms accounts for the observed modulation, it is possible to conjecture a model which does clarify the results. The incoming stream of particles (or the ionization they produce) is detected not by radio wave reflections from its earthward-approaching surface, but by reflections from the sides of the incoming stream. If the side surfaces of the stream are sufficiently irregular, reflections from the stream as these irregularities rush by may produce the observed Doppler effects. In general, the irregularities may be thought of as conical depressions or elevations whose angle with the surface of the stream is given by $\tan^{-1} x/y$. When the probing radio wave is directed perpendicular to the stream motion, then $y = 3 \times 10^8$ cm. A modulation of 200 c/sec

for example when probing at a radio frequency of 10 mc/sec could be produced if $x = 3 + 10^5$ cm. Compared to the size of the stream surface, the irregularities are relatively minute and non-cyclic, but they are nevertheless of the correct size to give rise to the rather

Table 2. Speed of the auroral reflecting curtain (cm/sec) towards the observer, necessary to produce the observed modulation

Observed Modulations (c/sec)	*Radio Probing Frequency (mc/sec)*		
	10	50	100
50	$7{\cdot}5 \times 10^4$	$1{\cdot}5 \times 10^4$	$7{\cdot}5 \times 10^3$
200	3×10^5	6×10^4	3×10^4

erratic fading records which are actually observed. The irregularities could easily be caused by local turbulence, diffusion, local changes in the rates of recombination or ionization, etc.

It should be noted that yet to be observed in the radio probings of the aurora are Doppler shifts of 0·2–2 mc/sec. These maximum effects would be found when the probing radio wave is directed along the path of the incoming protons.

Geophysics Research Directorate,
Air Force Cambridge Research Center. N. C. Gerson.

VOL. 4 1953

The electron content of the F_2 layer above Singapore

(*Received* 23 *June* 1953)

In this note we compare the total ionization present in a unit column, n, below the level of maximum ionization density of the F_2 layer in 1949, a year of great solar activity ($R \sim 120$), and in 1952, a year of slight activity ($R \sim 30$). The total electron content, n, is estimated by parabolic methods (Osborne, 1952*a*, Ratcliffe, 1951) for six equally spaced days in June and December 1952, and the results are compared with similar data for the corresponding months of 1949 (Figs. 1, 2).

The diurnal variations of "n" are reasonably consistent with the hypothesis that the ions are produced by photo-ionization in a medium with an appreciable ionization relaxation time. Perturbations due to movements of ionization, which dominate the corresponding

Values of "n" when $\chi = 23°$ (*cos* $\chi = 0{\cdot}92$)

Station	*Sunspot maximum*	*Sunspot minimum*	*Ratio*
Alaska/Watheroo	$1{\cdot}3 \cdot 10^{13}$	$0{\cdot}6 \cdot 10^{13}$	2
Huancayo	$2{\cdot}2 \cdot 10^{13}$	$1{\cdot}2 \cdot 10^{13}$ (estimated[3])	1·8
		$0{\cdot}8 \cdot 10^{13}$ (uncorrected)	2·8
Singapore—June	$2{\cdot}0 \cdot 10^{13}$	$1{\cdot}7 \cdot 10^{13}$	1·2
December	$2{\cdot}3 \cdot 10^{13}$	$1{\cdot}7 \cdot 10^{13}$	1·4

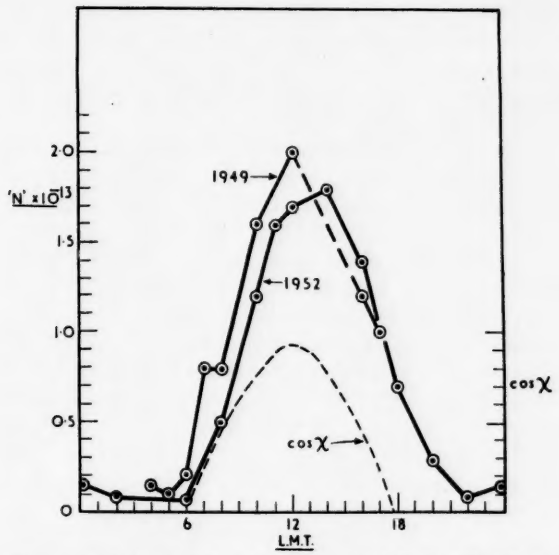


Fig. 1. The mean diurnal variations of the electron content "n" of the F_2 region over Singapore during June 1949 and June 1952.

curves for f_0F_2, modify the shape of the curves for 1949 but are comparatively small in 1952. Thus the large anomalous differences between June and December evident in sunspot maximum years have largely disappeared.

It is interesting to compare these results with those obtained by RATCLIFFE (1951*b*) for Huancayo, Alaska and Watheroo.

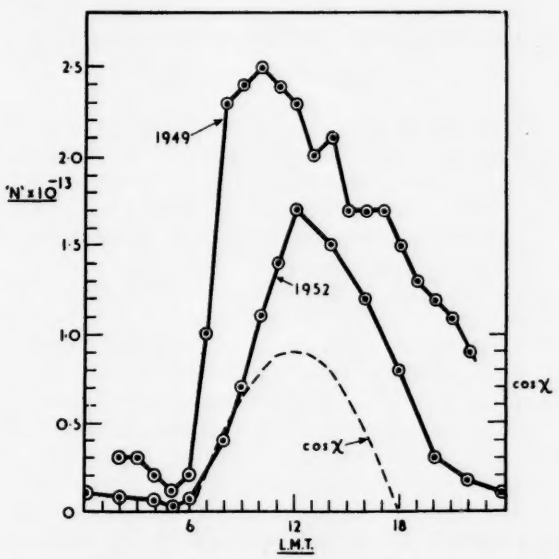


Fig. 2. The mean diurnal variations of the electron content "n" of the F_2 region over Singapore during December 1949 and December 1952.

The mid-day values of n at Singapore at sunspot maximum are very similar to those obtained by RATCLIFFE (1951) for Huancayo; and it is suggestive that the values of n at Singapore during the early morning ($\chi > 60°$) in 1952 are consistent with those found at Alaska and Watheroo for similar values of χ, again implying that transport of ionization is comparatively unimportant at these times.

The small change in "n" at Singapore between 1949 and 1952, and the differences between the shapes of the diurnal curves in 1949 and 1952, which are particularly noticeable during the December solstice (see Fig. 2), appear to be due (MCNISH, 1950; OSBORNE, 1952*b*) to ionization moving above the level of maximum ionization density in years of great solar activity. RATCLIFFE (1951*b*) however claims that this phenomenon is most evident at Huancayo in years of minimum solar activity, though MCNISH has observed its presence there during sunspot maximum (1939) also.

The vertical incidence ionospheric records used for this analysis were obtained at the Singapore Radio Research Substation of the Department of Scientific and Industrial Research. This note is published by permission of the Director of Radio Research of that Department.

University of Malaya,
Singapore.

B. W. OSBORNE.

REFERENCES

MCNISH, A. G.	1950	Proc. Conf. on Ionospheric Physics, Pennsylvania State College, Section CC, pp. 9, 10
OSBORNE, B. W.	1952*a*	J. Atmosph. Terr. Phys., **3**, 58
	1952*b*	Nature, Lond. **169**, 661
RATCLIFFE, J. A.	1951*a*	J. Geophys. Res. **56**, 463
	1951*b*	J. Geophys. Res. **56**, 487

VOL.
4
195

Induced charges and the fine-weather field

(*Received* 17 *July* 1953)

When WHIPPLE (1929) worked out the suggestion of APPLETON (1925) that the variation with time of day of the earth's vertical field in fine-weather "undisturbed" conditions is caused by the variation with time of day of the total thunderstorm activity over the whole earth, he considered that the charge on the fine-weather portion of the earth, and hence the vertical field as measured, can be altered by (1) fine-weather conduction current and (2) currents to the earth during storms.

There is, however, a third factor, namely a variation in the amount of charge induced on the earth's surface below clouds; if a storm cloud were to be considered as simply a dipole above a conducting earth, then the two charges induced on the earth would be equal and opposite, but this does not take into account induced charges on the ionosphere nor possible inequalities in the two quantities of charge in the cloud.

If $-\sigma$ is the fine-weather surface density of charge, the fine-weather field $F = 4\pi\sigma$; then the total charge on the fine-weather portions of the earth is $-Q = -A\sigma$ where A is the area of the earth's surface (storm areas form a sufficiently small fraction to be neglected here). If λ is the conductivity of the atmosphere close to the earth, the fine weather conduction current is $F\lambda$ per unit area, giving a total of $4\pi\lambda Q$. Let us suppose that, at any instant, there are N "average" storms in action, each providing a current $-j$ to the earth and each inducing a charge q on the earth.

Then, for the whole earth:

$$-\frac{dQ}{dt} = 4\pi\lambda Q - Nj - q\frac{dN}{dt} \qquad (1)$$

WHIPPLE's (1929) equation omits the last term.

If we assume that the number of storms in action varies with a 24 hour period according to:

$$N = X + Y\cos wt$$

it is easy to show that:

$$Q = B + C\cos(wt + \alpha)$$

where

$$B = \frac{jX}{4\pi\lambda}\cdot C = Y\left[\frac{j^2 + q^2w^2}{w^2 + (4\pi\lambda)^2}\right]^{\frac{1}{2}}$$

and

$$\tan\alpha = \frac{(j - 4\pi\lambda q)w}{4\pi\lambda j + qw^2}$$

WHIPPLE's results, leaving out the term in q, correspond to $\tan\alpha = w/4\pi\lambda$, giving a time lag between the maxima of N and Q which he calculated to be about 12 minutes, quite inappreciable on his curves.

The difference from WHIPPLE's results depends on the ratio of q to $j/4\pi\lambda$. From the results of GISH and WAIT (1950), j can be taken as 0·5 amp. MAUCHLY (1926), at sea, found an average value of λ of $2{\cdot}3 \times 10^{-4}$ E.S.U. (which would give a time lag $1/4\pi\lambda = 5{\cdot}8$ minutes), so that $j/4\pi\lambda = 174$ coulombs. Estimates of q are difficult to obtain; GISH and WAIT (1950) for a "less electrically active" cloud found that the fields above the cloud could be accounted for by charges in the cloud of $\pm$ 39 coulombs, but the charge induced on the earth would be considerably less than this. Other estimates of charges in clouds, by SIMPSON and ROBINSON (1940) and by MALAN (1952) are of the same order of magnitude.

Thus it seems highly probable that q is smaller than $j/4\pi\lambda$, perhaps of the order of 10%, so that the additional term introduced into the equation is not of much practical importance; but it has needed the present analysis to show this.

The correlation between field maximum and thunderstorm maximum has been taken as a proof that storms bring negative charge to the earth. But equation (1) shows that if λ and j were both small, Q would vary with N, due to the effect of induced charges rather than of charges brought down by the storm; and measurements of λ and j are required in addition to the correlation of maxima to give a definite conclusion.

Durham Colleges,
University of Durham.

J. A. CHALMERS.

REFERENCES

APPLETON, E. V.	1925	Proc. Phys. Soc. **37,** 49
GISH, O. H. and WAIT, G. R.	1950	J. Geophys. Res. **55,** 473
MALAN, D. J.	1952	Ann. de Geophys. **8,** 385
MAUCHLY, S. J.	1926	Carnegie Inst. Res. Dept. Terr. Mag **5,** 385
SIMPSON, G. C. and ROBINSON, C. D.	1940	Proc. Roy. Soc. A. **177,** 281
WHIPPLE, F. J. W.	1929	Q.J. Roy. Met. Soc. **55,** 351

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 86 to 89. Pergamon Press Ltd. London

Indices of geomagnetic activity

of the Observatories ABINGER (*Ab*), ESKDALEMUIR (*Es*) and LERWICK (*Le*)

April 1953 to August 1953

The figures given on pages 86 to 89 represent the K-indices for three-hour intervals, beginning with 00–03 hrs for the first and ending with 21–24 hrs for the eighth figure.

April 1953

Day	*Ab* Range for $K = 9 : 500\gamma$		*Es* Range for $K = 9 : 750\gamma$		*Le* Range for $K = 9 : 1000\gamma$		*Day*
	K-Indices	*Sum*	*K-Indices*	*Sum*	*K-Indices*	*Sum*	
1	3123 3322	19	3123 3212	17	4113 2221	16	1
2	4332 2343	24	3222 2233	19	4321 1233	19	2
3	4222 2343	22	3211 1244	18	3311 1247	22	3
4	4433 3344	28	5422 3344	27	8522 3343	30	4
5	2022 2222	14	2012 2322	14	1011 1332	12	5
6	2322 2123	17	2222 1112	13	2221 1122	13	6
7	0122 1222	12	0001 2112	7	0000 1112	5	7
8	2223 2214	18	2121 2223	15	3211 1213	14	8
9	3311 2222	16	3211 2221	14	3211 2211	13	9
10	2323 3334	23	2223 3323	20	1222 3324	19	10
11	1433 3533	25	1333 3532	23	1333 3432	22	11
12	3333 3233	23	3223 3223	20	3232 3132	19	12
13	4233 4431	24	3233 4331	22	3323 3331	21	13
14	2222 3122	16	2221 2212	14	2111 2221	12	14
15	3323 3232	21	3223 3222	19	2222 2221	15	15
16	4444 3556	35	3433 3455	30	4533 4555	34	16
17	3223 3312	19	3223 3212	18	5322 3211	19	17
18	2233 3332	21	2132 2322	17	2222 2222	16	18
19	3422 4424	25	2422 4324	23	2433 3325	25	19
20	4444 4435	32	4333 4434	28	4432 3334	26	20
21	3343 3544	29	3343 3444	28	4333 3444	28	21
22	3333 4443	27	2323 4443	25	3322 3334	23	22
23	3432 3344	26	3432 3343	25	4542 3343	28	23
24	2111 3343	18	2012 2332	15	3111 2322	15	24
25	3223 3334	23	2112 3333	18	1112 2334	17	25
26	2233 4314	22	2232 4324	22	3231 3313	19	26
27	3123 2223	18	3222 3222	18	3312 3223	19	27
28	2233 2310	16	2223 2210	14	2222 1110	11	28
29	1123 4310	15	1123 4321	17	1122 3210	12	29
30	3234 3442	25	2223 3432	21	2222 2431	18	30

VOL. 4 1953

May 1953

Day	*Ab* Range for $K = 9 : 500\gamma$		Sum	*Es* Range for $K = 9 : 750\gamma$		Sum	*Le* Range for $K = 9 : 1000\gamma$		Sum	Day
	K-Indices			*K-Indices*			*K-Indices*			
1	1122	3310	13	1121	2311	12	1011	2311	10	1
2	2122	3111	13	2102	2111	10	2102	2010	8	2
3	0011	1322	10	0111	1222	10	1111	1222	11	3
4	1312	1211	12	1212	1201	10	1111	1201	8	4
5	2323	2333	21	2223	2233	19	1322	2223	17	5
6	4345	4455	34	4334	4445	31	4333	3345	28	6
7	4444	3333	28	4344	3323	26	4343	3213	23	7
8	4343	4344	29	4343	4343	28	4342	4343	27	8
9	4333	3333	25	4333	3333	25	4333	2334	25	9
10	2123	4332	20	2112	4332	18	3112	3212	15	10
11	2123	2221	15	2122	2211	13	2122	2211	13	11
12	2211	1222	13	2120	2222	13	1120	2111	9	12
13	1122	3111	12	1012	2221	11	0012	2111	8	13
14	1111	2212	11	1111	2222	12	1011	1112	8	14
15	1323	5566	31	1323	4566	30	1322	4487	31	15
16	4534	5566	38	4544	5556	38	6554	6677	46	16
17	6432	3434	29	5422	3333	25	7531	2324	27	17
18	4433	3333	26	3333	3333	24	3332	3234	23	18
19	4434	4244	29	3433	4234	26	4433	3134	25	19
20	3233	3333	23	3233	3333	23	4323	2333	23	20
21	2332	2333	21	2232	2332	19	2221	1332	16	21
22	2322	3343	22	2323	3343	23	2412	3243	21	22
23	3113	2232	17	3122	3332	19	3112	2221	14	23
24	1121	3332	16	1121	3332	16	1121	3221	13	24
25	0123	2323	16	1122	2312	14	0121	2203	11	25
26	1112	3243	17	2102	2233	15	2112	2132	14	26
27	3444	3432	27	3444	3322	25	2433	3322	22	27
28	2322	2331	18	2322	2321	17	2212	2321	15	28
29	1122	3211	13	1122	3212	14	1122	3211	13	29
30	2221	2222	15	2110	2222	12	3211	2212	14	30
31	2113	3342	19	2002	3342	16	1111	3341	15	31

June 1953

Day	*K-Indices*		Sum	*K-Indices*		Sum	*K-Indices*		Sum	Day
1	2321	1232	16	2221	2221	14	1211	1221	11	1
2	2455	5554	35	2454	4454	32	3443	4464	32	2
3	4354	4443	31	4343	4433	28	6343	3433	29	3
4	4444	3432	28	4433	3422	25	4443	3322	25	4
5	1343	2352	23	1333	2352	22	2332	2352	22	5
6	2433	3331	22	2333	3332	22	3322	2321	18	6
7	2232	3321	18	2232	2222	17	2232	2211	15	7
8	1221	2221	13	1121	2221	12	1121	2111	10	8
9	1222	2221	14	1122	2221	13	1112	2210	10	9
10	1333	3342	22	2323	3332	21	1322	2232	17	10
11	2222	3324	20	2222	3223	18	2212	2113	14	11
12	1223	3444	23	1123	3444	22	1112	2333	16	12

June 1953 (*contd.*)

Day	*Ab* Range for $K = 9 : 500\gamma$		Sum	*Es* Range for $K = 9 : 750\gamma$		Sum	*Le* Range for $K = 9 : 1000\gamma$		Sum	Day
	K-Indices			*K-Indices*			*K-Indices*			
13	3323	3343	24	3322	3333	22	4411	2233	20	13
14	2222	3332	19	2222	3332	19	2111	3231	14	14
15	0122	3232	15	1121	2232	14	1111	1121	9	15
16	1113	3222	15	1112	3212	13	0112	2111	9	16
17	1122	4423	19	1222	4422	19	1121	3312	14	17
18	2232	1222	16	2222	1122	14	2221	2111	12	18
19	2121	2321	14	1221	2221	13	1211	2211	11	19
20	1233	4443	24	1123	3343	20	1122	3343	19	20
21	2333	2333	22	2233	2333	21	2332	2233	20	21
22	1122	4333	19	1222	3333	19	1222	3312	16	22
23	1111	2112	10	2111	1102	9	1111	1111	8	23
24	2222	2333	19	2222	2233	18	1211	1232	13	24
25	2122	2221	14	2111	2221	12	2111	1121	10	25
26	0012	1321	10	1012	1221	10	0010	1121	6	26
27	1113	2322	15	1122	2222	14	1112	1222	12	27
28	1212	2123	14	0211	1023	10	0111	1012	7	28
29	1135	6644	30	2134	6544	29	1134	5544	27	29
30	5455	4443	34	5344	4443	31	6434	3333	29	30

July 1953

Day	*K-Indices*		Sum	*K-Indices*		Sum	*K-Indices*		Sum	Day
1	3344	4333	27	3344	4333	27	3334	4223	24	1
2	4443	4444	31	4443	4444	31	5443	4333	29	2
3	3233	4433	25	3223	4442	24	3322	3333	22	3
4	3334	4442	27	3334	4342	26	3333	3331	22	4
5	3333	3332	23	2232	2322	18	2232	2223	18	5
6	4222	3323	21	4222	2223	19	5212	2212	17	6
7	3332	4444	27	3432	4434	27	3442	3333	25	7
8	4223	3343	24	3223	2343	22	4222	2243	21	8
9	3123	3343	22	2122	3343	20	2121	2233	16	9
10	3223	2211	16	2222	1210	12	3221	1201	12	10
11	2322	2121	15	2322	2211	15	1311	2011	10	11
12	1122	3354	21	1212	2344	19	1111	2254	17	12
13	3233	3332	22	2232	3322	19	2232	3222	18	13
14	1112	3542	19	1012	3442	17	1111	2331	13	14
15	3333	4431	24	3333	3331	22	3332	3321	20	15
16	1222	1211	12	1212	0112	10	1111	1111	8	16
17	1123	3211	14	1122	2211	12	1112	2110	9	17
18	2212	3212	15	2212	2212	14	2211	2111	11	18
19	3322	3122	18	3322	1122	16	3211	0112	11	19
20	2222	3311	16	2212	3312	16	2211	3301	13	20
21	2123	3213	17	2122	2213	15	2112	1102	10	21
22	0122	1333	15	1123	1323	16	0012	1213	10	22
23	3434	5545	33	3434	5545	33	3434	5545	33	23
24	4223	3433	24	4213	3333	22	3112	3322	17	24

VOL. 4 1953

July 1953 (*contd.*)

Day	*Ab* Range for $K = 9 : 500\gamma$		Sum	*Es* Range for $K = 9 : 750\gamma$		Sum	*Le* Range for $K = 9 : 1000\gamma$		Sum	Day
	K-Indices			*K-Indices*			*K-Indices*			
25	1122	3442	19	0122	3442	18	0121	2332	14	25
26	4333	4343	27	4322	3343	24	5322	3234	24	26
27	4553	5543	34	4543	4543	32	4543	4443	31	27
28	4443	3444	30	3443	3443	28	4433	3333	28	28
29	4444	3454	32	4433	3454	30	4443	3344	29	29
30	4343	4444	30	4233	3444	27	5333	3334	27	30
31	4334	4331	25	4333	3322	23	4433	3221	22	31

August 1953

Day	*K-Indices*		Sum	*K-Indices*		Sum	*K-Indices*		Sum	Day
1	3323	3444	26	2322	3334	22	3312	3333	21	1
2	3323	2333	22	2333	2332	21	2322	2322	18	2
3	3122	3333	20	3222	3322	19	2121	3221	14	3
4	3212	2332	18	2222	2322	17	3321	2321	17	4
5	2213	2133	17	2122	2232	16	1111	1132	11	5
6	0123	3432	18	1012	2422	14	1111	2322	13	6
7	3323	3333	23	3212	2222	16	3312	1222	16	7
8	3223	2212	17	3222	1222	16	3211	1211	12	8
9	3323	4353	26	3322	4342	23	4321	3343	23	9
10	3343	2434	26	2243	2334	23	2243	3325	24	10
11	4433	3443	28	4322	3432	23	5432	3333	26	11
12	4544	4543	33	4544	4543	33	5554	4544	36	12
13	4332	4455	30	4321	4455	28	5431	3355	29	13
14	3322	2332	20	3322	2332	20	3312	2321	17	14
15	2122	2243	18	2122	2243	18	1111	2243	15	15
16	3134	3332	22	3123	3332	20	3123	2322	18	16
17	2112	3221	14	2121	3311	14	1111	2200	8	17
18	3222	2322	18	2212	2322	16	3211	2211	13	18
19	2112	3221	14	1112	2211	11	1101	1211	8	19
20	2112	2211	12	2112	2112	12	1010	1001	4	20
21	1132	2221	14	1112	2210	10	1101	1111	7	21
22	0222	3111	12	0112	2211	10	0101	2222	10	22
23	4335	4556	35	4324	4545	31	3324	4456	31	23
24	5534	4655	37	5433	4655	35	6433	3457	35	24
25	3345	4333	28	4344	4333	28	5434	4333	29	25
26	4433	3445	30	4322	3435	26	4422	3334	25	26
27	4344	4554	33	4434	4554	33	5523	4544	32	27
28	4334	4545	32	4434	3444	30	4433	3445	30	28
29	3434	3554	31	3444	3453	30	4443	3453	30	29
30	4444	5544	34	5444	4544	34	6443	4434	32	30
31	4424	4444	30	4323	4443	27	4423	4343	27	31

Book review

GILLESPIE, A. B; **Signal, Noise and Resolution in Nuclear Counter Amplifiers.** (Vol. 1 of Electronics and Waves—A Series of Monographs. General Editor: D. W. Fry, A.E.R.E., Harwell.) Demy 8vo, xii + 155 pp., 61 diagrams. Pergamon Press Ltd., 1953. (Published in U.S.A. by McGraw-Hill, Inc., New York.) 21/- net or $4.50.

The science and art of electronics have expanded greatly under the stimulus of the war, and the new tools are being handled, somewhat timidly, by scientists in other fields. The timidity comes in part from the unfamiliar concepts which electronics has created for itself, and from the correspondingly unfamiliar jargon of its devotees. A straight textbook on an electronic subject is usually quite hard reading for the medical man or nuclear physicist who may want to improve or even design his own electronic apparatus.

Pergamon Press are to be congratulated on having started this series of Monographs on Electronics and Waves, and on having acquired D. W. FRY as an editor. If the first volume is a representative sample the series ought to prove very useful indeed, acting as the much-needed interpreter between the communication engineer and a variety of other scientists.

Amplifiers for the detection of nuclear particles came into use as early as 1924, but for some twenty years their construction was an art rather than a science. Some attempts to analyse the problems and to design amplifiers for optimum performance were made, but it was only when communication engineers started to take a hand that real progress resulted. That in turn persuaded some people that the time had come to leave the design of "nuclear" amplifiers to electronic specialists, and that soon the nuclear physicist would buy his electronic apparatus all ready made like a doctor buys his microscope. But it turned out that the problems of nuclear physics were too variable and that success often depended on the ability to tailor one's equipment to fit a given problem. And for that it is essential to master the interrelations between the various desirable qualities of a nuclear amplifier as well as the means for getting the best compromise between their conflicting claims.

A. B. GILLESPIE concentrates on the problem of getting the best signal-noise ratio under the conditions set by a given experimental problem. He goes over the ground very carefully and his explanations are clear and explicit throughout. Mathematical derivations are banished into a number of appendices; the text contains only the formulae needed for actual computations, and brief indications how they are derived. As a result the book makes easy and pleasant reading. It also contains a number of practical remarks on the choice of valves, operation conditions etc. I think it will be found indispensible by any nuclear physicist who wishes to design his amplifiers.

Perhaps we may hope, in the same series, for a companion volume dealing with the later stages of amplification, with problems of stability and overloading, with stabilised voltage supplies, and with scaling and sorting devices.

O. R. FRISCH.

BIRKS, J. B.; **Scintillation Counters.** (Vol. 2 of Electronics and Waves—A Series of Monographs. General Editor: D. W. Fry, A.E.R.E., Harwell.) Pergamon Press, London, 21s. and McGraw-Hill, NewYork, $4.50, 1953. 148 pp.

CURRAN, S. C.; **Luminescence and the Scintillation Counter.** Butterworths Scientific Publications, London, 1953. 32s. 6d. 219 pp.

The technique of scintillation counting has gained rapidly in importance during the last few years, and has indeed in many applications superseded such well established instruments as the Geiger counter or the proportional counter. A book which would summarise the relevant information was urgently needed, and now two such books have appeared in quick succession.

It is not easy to say which of the two is the better. Both authors are very well qualified for their task. CURRAN is an authority on nuclear physics and was one of the early pioneers of the scintillation method. His book contains a great wealth of information, and is on the whole the more ambitious one. It not only describes photomultipliers and scintillating materials, but also goes deeply into the physical phenomena which enter in their use. A whole chapter is

VOL. 4 1953

devoted to the interaction of various radiations with matter, and another one to the phenomenon of secondary emission.

BIRKS is known for important contributions to the mechanism of luminescence. His book is somewhat slighter and less ambitious, but it might well be found to be the more useful one by those who merely want to use scintillation counters without going too deeply into their fundamental principles. It contains all the necessary information for that purpose presented in a straightforward and very readable style. This is another volume in the series Electronics and Waves, and the Pergamon Press are to be congratulated on having added such an excellent presentation of a topical subject at such unusual speed (the book appeared within two months of delivery of the manuscript).

O. R. FRISCH,

Advances in Electronics, Vol IV, Edited by L. MARTON. pp. 344. Academic Press, Inc., N.Y., 1952. $7.80.

The title of this book will seem somewhat misleading to readers who are not already familiar with the series of which it forms the fourth volume. The term electronics has come by force of common usage, to mean that branch of applied science associated with the use of valves, electron tubes and the like; techniques in which circuitry and devices exploiting the free electron in vacuo predominate. It is therefore with a shock of surprise that one observes that the first article in the present volume is a review by Professor H. S. W. MASSEY of electron scattering in solids. The article is what one expects from Professor MASSEY, an authoritative survey, mainly from the theoretical standpoint, of many phenomena that are best described or explained in terms of the scattering of electrons by a close-packed assemblage of atoms constituting a solid. Secondary emission and electron diffraction are excluded as being adequately treated elsewhere, but there is a full discussion, with 87 references, of single elastic scattering of fast electrons, of single inelastic scattering such as is relevant to the electron microscope, of multiple scattering and diffusion and the bearing of the various processes on the resistivities of pure metals, alloys and semi-conductors. All this is a valuable contribution to scientific literature but one cannot help wondering whether it would not have been more effective if published within different covers.

The remaining articles in the volume come much nearer to what is ordinarily understood by the term electronics. The only other theoretical article is one by A. VAN DE ZIEL on Fluctuation Phenomena, a review providing the electronics practitioner with a comparatively simple treatment by Fourier analysis of a number of important problems associated with what he experiences as noise; thermal noise in generators, shot noise in diodes and triodes. On the practical side there is a lucid account by G. A. MORTON of the scintillation counter and its component parts, the photomultiplier and the phosphor crystal, together with a review of their applications in nuclear physics; also detailed surveys of the modulation of continuous wave magnetrons by J. S. DONAL, Jr. and of the technique of magnetic surveying from aircraft, especially with the saturable core magnetometer, as used for submarine detection during the war and for geophysical work since. An article by C. V. L. SMITH gives a general account of the position in the field of electronic digital computors at the end of 1950, when *Whirlwind* and SEAC had, in the author's words, "come to fruition" but many machines that have since done so were still in the budding stage. The final article by M. G. PAWLEY and W. E. TRIEST gives a short general account of multichannel radio telemetering or systems for making measurements of several variables at a distance with radio transmission as the connecting link between the observer and the locality to which the measured quantities refer. The importance of such techniques in providing data for localities not accessible to any human observer is obvious enough.

It will be clear from what has been said that the book provides a very useful group of surveys covering particular fields of topical interest. Many will be glad to find it in the library but probably few will want it on the personal bookshelf.

L. HARTSHORN.

DEUTSCH, S.; **Theory and Design of Television Receivers.** xix + 536 pp. 386 illustrations. McGraw-Hill Book Company, Inc. 1951. Price **$6.50.**

While the literature relating to the scientific and technical development of television is already vast, there is a dearth of books dealing adequately with the basic principles of the subject in a manner appropriate to the professional radio or electrical engineer. The book under review aims at filling this gap by presenting a physical explanation of the behaviour of various television receiver circuits, developing the mathematical theory of these circuits and discussing the practical design of receivers.

The first chapter on "Television Standards" makes it clear that the work is to be confined in its quantitative aspects to the American practice of using a scanning rate of 525 lines on pictures presented 30 times a second. The reader will have no difficulty, however, in converting the numerous examples and problems discussed to the standards in use in the United Kingdom (405 lines, 25 pictures a second), or elsewhere. Chapters 2 and 3 are devoted to a mathematical discussion of the characteristics of amplifier and rectifier circuits, and the properties of the associated tuned coupling circuits. Succeeding chapters then deal with each of the basic sections of a receiver from the antenna to the picture tube, including power supplies and frequency control. A discussion on direct- and projection-viewing systems follows; and the concluding chapter illustrates the various faults and misalignments that may occur in a television receiver, and gives brief details of the service tests required to detect and rectify these.

Each chapter is followed by a series of problems or exercises, the solution of which should enable the reader to gain a more precise and detailed understanding of the subject. The production of the book, which is liberally illustrated, is excellent, and the work may be recommended to all those studying the subject of television reception.

R. L. SMITH-ROSE

VOL.
4
1953

CONFERENCE ON ROCKET RESEARCH OF THE UPPER ATMOSPHERE

The Conference which was held in Oxford, August 24 to 26 1953, under the auspices of the Gassiot Committee of the Royal Society, was concerned with problems of the physics of the upper atmosphere of the earth, with particular reference to the results obtained by the techniques of rocket research. Topics discussed included measurements of atmospheric temperatures and particle densities, measurements of the electron density distribution in the ionized layers, of the changes in chemical composition due to diffusive separation and of the earth's magnetic field and cosmic ray activity at high altitudes.

It is intended to publish the proceedings of the Conference as a separate supplement volume to the *Journal of Atmospheric and Terrestrial Physics*. The volume can also be purchased separately through the book trade.

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 93 to 105. Pergamon Press Ltd., London

Mobility of atmospheric small ions during summer nights at Uppsala

H. Norinder and R. Siksna

Institute for High Tension Research, University of Uppsala, Sweden

(*Received* 4 *July* 1953)

ABSTRACT

A method involving the periodic alternation of limiting mobilities of a simple aspiration condenser has been employed to obtain characteristics of the number of atmospheric small ions captured during suitable weather conditions. The distributions of ions with respect to mobilities were computed. The structure of the distribution function varied with time. The idea is discussed that ions which are commonly called atmospheric small ions consist of two groups: small ions proper or cluster ions with mobilities between 1·5–2·5 cm sec^{-1}/volt cm^{-1}, and affected small ions with mobilities between (0·3 ÷ 0·5) — (1 ÷ 1·5) cm sec^{-1}/volt cm^{-1}. Confusions which have hitherto existed concerning the mobilities of atmospheric small ions could be obviated by this conception. It is therefore of interest to investigate atmospheric small ions under the conditions which exist in the atmosphere, i.e. together with the factors which affect the pure small ions. Results obtained up till now have been encouraging, in spite of measuring difficulties caused by the irregular variations of the ionic concentrations investigated.

1. Introduction

The ions present in the atmospheric air are of interest from many points of view. The ionic state of the atmospheric air is often given as the number of ions per cm^3, or, in most cases, the concentration is given together with an indication of the kind of the ions concerned—small ions, intermediate ions, large ions. But our knowledge of the properties of the atmospheric ions themselves is slight. The mobility of the ions is one of the quantities that may be measured, and it provides a characteristic property of the ions present. But measurement of the mobility of atmospheric ions is associated with some difficulties; the gustiness of the variations in the concentration of ions in the atmospheric air is one of the main difficulties encountered in such measurements.

The simplest and the most used method for determining the mobility of atmospheric small-ions is the simultaneous measurements of the conductivity λ, and the ionic density n. Assuming that the ions present have only one discrete mobility k, this may be computed from the relation

$$\lambda = nke,$$

where e is the elementary charge. This relation has been used recently: for computing the mobility of small-ions from measurements made during the last cruise of the ship *Carnegie* (Torreson *et al.*, 1949) for computations on the results of measurements carried out during 1935–37 at Poona (India) (Agarwala, 1951), and for computing the average mobility from measurements obtained in New York City and in its surroundings (O'Donnel, 1952). It is evident that the mobility obtained in this way is a mean mobility, because ions of different mobilities, present in different concentrations, contribute to the conductivity. This mean mobility is defined by the formula

$$\bar{k} = \frac{\Sigma n_i k_i}{\Sigma n_i},$$

where n_i is the concentration of ions with a mobility k_i. The same mean mobilities may be obtained by using the direct aspiration methods of GERDIEN (1903) and MACHE (1903).

The data on the mobilities of atmospheric small ions surveyed in the monographs on the subject (HESS, 1926) have been obtained with the above-mentioned indirect and direct methods, and they must be regarded as data on the mean mobilities.

Recently, FEDOROW (1952) has described a method which, as indicated by the author, might be developed into a recording method. It is possible by this method to determine the mobility spectrum of atmospheric ions in 2–3 minutes by measuring the ionic current in an ionization chamber into which a sample of the air to be examined is aspirated. VASSAILS (1949) had earlier succeeded in investigating large laboratory ions by a similar method. The method seems to be suitable for investigation of the larger ions, but for examination of small atmospheric ions difficulties might arise because of transformation of the small ions present when left for even 2–3 minutes in a closed ionization chamber.

Encouraged with the results of examination of the mobility distribution of laboratory ions of different kinds (NORINDER and SIKSNA, 1952*b*; SIKSNA, 1952*a*, 1952*b*, 1953) obtained by using the method of the periodical alternation of chosen limiting mobilities of an aspiration ionic counter, we have attempted to use the same method for investigating atmospheric small ions. It was evident that the method could not be used in all atmospheric conditions, because of the gusty variation of ionic concentration in various circumstances. It seemed to us that suitable conditions for such measurements could be expected during quiet summer nights at Uppsala, when, as has been shown, the number of ions increases because of accumulation of radioactive substances in the lower layers of the atmosphere (NORINDER and SIKSNA, 1950, 1952*a*). We have also attempted to use the same method under other conditions. We did not succeed in obtaining computable results when there were S, SW winds, because the number of ions was low during these weather conditions, and liable to sudden variations. When there were NW winds the ionic conditions were comparatively constant, and computable results were obtained from the measurements. In the following we will describe the results of measurements of atmospheric small ions during quiet summer nights at Uppsala.

VOL.
4
195

2. THE METHOD EMPLOYED

As already mentioned, the distribution of atmospheric small ions with respect to mobilities was determined by the method of a periodical alternation of chosen limiting mobilities of a cylindric aspiration ion-counter. The method gave the time curves of the number of those ions captured at the chosen limiting mobilities. Conclusions concerning the distribution of the ions present with respect to mobilities may be drawn from the shape of the curves. For this purpose, the curves must be intersected by vertical straight lines at particular instants and the numbers obtained may be plotted in $n = F(1/k_*)$ characteristics (number of ions n as a function of the reciprocal adjusted limiting mobility k_*), and from these curves a general

survey of the mobility distribution may be obtained. The detailed distribution function $f(k)$ may be obtained from the characteristics by the following calculation.

The number of ions was measured with an ion-counter for small ions which incorporated a Weger aspiration condenser and was built at the Institute (SIKSNA,

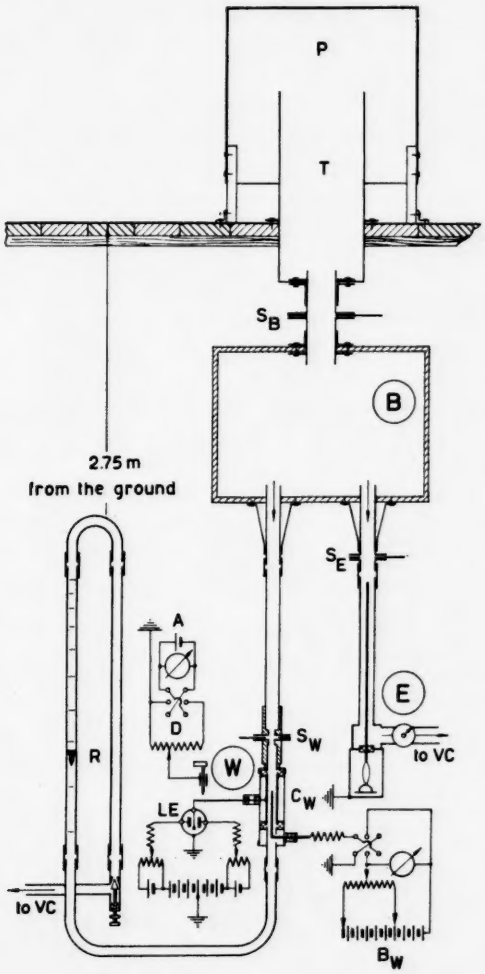


Fig. 1. The apparatus used: (*W*) Small ion counter with Weger condenser C_W, B_W battery for the condenser, *LE* Lindemann electrometer, *D* calibration device for the electrometer, *R* rotameter, S_W slide for the condenser, *VC* vacuum cleaner; (*E*) Ebert ion-counter, S_E slide for the Ebert counter; (*B*) suction box, S_B slide of the suction box, *T* connection tube to the open air, *P* protection box.

1953; WEGER, 1935). The arrangement of this small-ion counter and the auxiliary components is shown in Fig. 1. A detailed description of the counter will be given later. The counter has some good properties which seem to make it a suitable simple apparatus for multiple measurements of small-ions. For instance, it is

without the troublesome entrance end-effect. Moreover, it does not involve the polarization effect of the insulators when the potential of the central electrode is changed when adjusting the different limit mobilities. Characteristics of the small ions present were obtained with the ion-counter by varying the potential applied

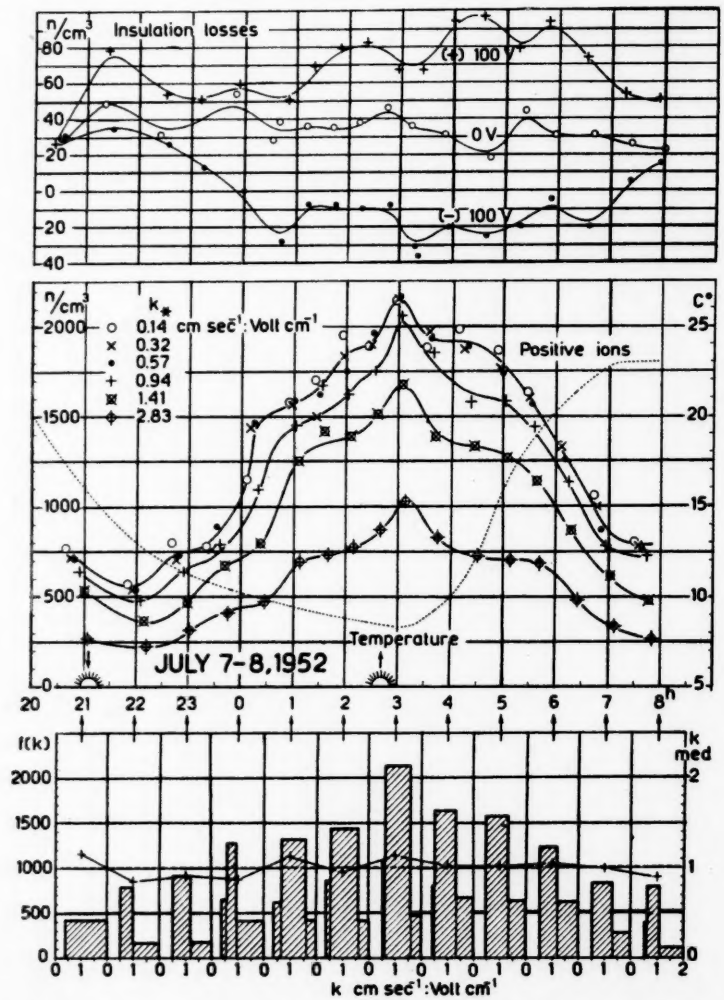


Fig. 2. Insulation losses; variation of the number of *positive* ions counted at chosen limit mobilities k_* of an aspiration condenser; $f(k)$ mobility distribution of ions present; median mobility k(med).

to the inner electrode of the Weger condenser, the air-flow through it remaining constant. The air to be examined was sucked into the counter from a suction box attached to the ceiling of the observation hut where the measurements were carried out. Besides the ion-counter with the Weger condenser, an Ebert counter was connected with the suction box—this counter being used to maintain a

VOL. 4 1953

sufficient air-flow through the suction box, because the aspiration through the small-ion counter alone seemed to be too slow to replace sufficiently rapidly the air in the suction box. The outside air was brought into the suction box through a tube which ascended through the roof of the hut. A protecting box was placed over this tube. The observation hut was located at the same place where earlier measurements of atmospheric ions have been carried out at this Institute (NORINDER and SIKSNA, 1952*a*)—57 m NW of the Institute main building. The measurements were carried out by alternating the potential applied to the inner electrode of the Weger condenser. At each potential, measurements were carried out for both the positive and negative values, giving the numbers of positive and negative ions captured. Prior to and after each series of measurements including all the chosen limiting mobilities, the insulation losses were measured by stopping the air-flow through the counter and closing the connection with the suction box. The insulation was measured alternately at both the polarities of the potential applied to the inner electrode of the Weger condenser. The insulation losses were not very great, but they varied with time, increasing when the content of radioactive substances in the air could be expected to be higher. The insulation losses also varied with the polarity of the potential applied to the inner electrode (Figs. 2 and 4). This phenomenon may be explained by the different shapes of the electrodes in the Weger condenser, which produce different conditions for deposition of the radioactive substances from the air sucked through the condenser. For several series, the insulation losses were also measured with the inner electrode earthed. In this case the values obtained were between those at the positive and negative potentials (Fig. 2).

3. Results of Measurements

Suitable weather conditions were not frequent during the summer of 1952. In some cases when the conditions seemed to be suitable at the beginning of the measurements, they changed during the night and only fragmentary series were obtained.

Two characteristic complete series of measurements will be considered in the following.

On July 7–8, 1952

The *weather conditions* were the most suitable for the measurements. On July 7 a region of very high pressure situated over Scandinavia shifted to the NE. Warm fine weather prevailed almost all over the country. On July 8 the high pressure region over Scandinavia weakened, but the dry and sometimes very warm weather was expected to last some days.

Results of the measurements are shown in Figs. 2 and 3. In Fig. 2 the insulation losses are shown in effective numbers of ions. As already mentioned, these curves characterize in a certain way the concentration of radioactive substances in the air.

The curves showing the number of ions

The numbers of ions at the three lower adjusted limit mobilities $k_* = 0{\cdot}14$, $0{\cdot}32$, $0{\cdot}57$ coincide, except for some individual points and during the decrease of

the concentration of negative ions after sunrise. This fact indicates that the lower k_* values used were in the region of saturation.

Mobility distribution of ions present

The intersections of the curves giving the number of ions by vertical straight lines at hourly intervals provided values from which the instantaneous characteristics

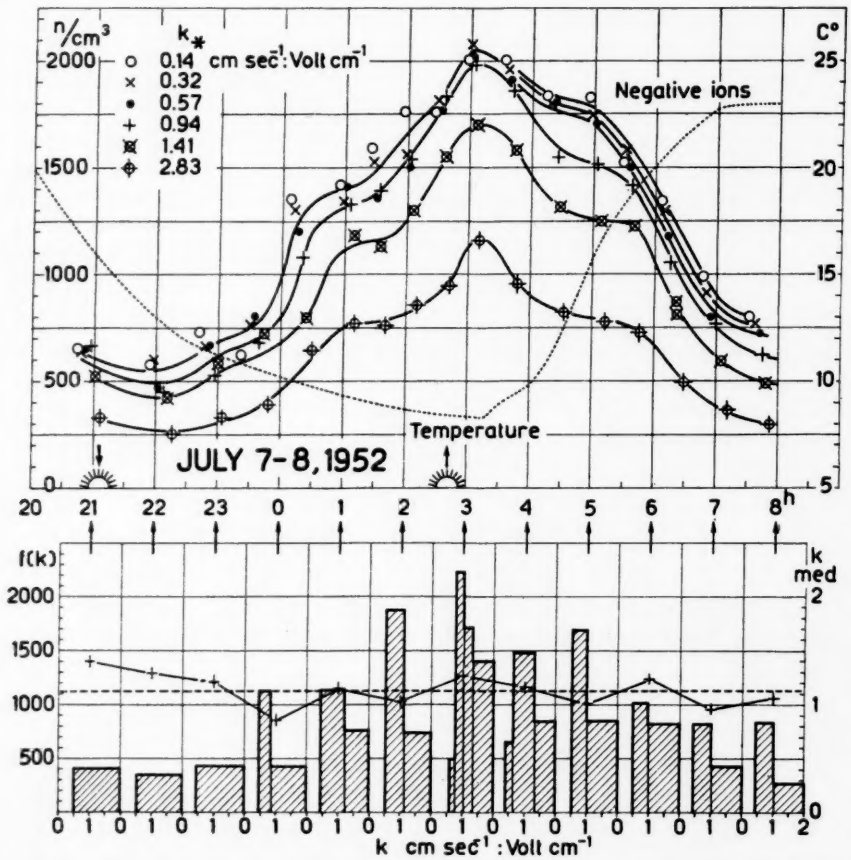


Fig. 3. Variation of the number of *negative* ions counted at chosen limit mobilities k_* of an aspiration condenser; $f(k)$ mobility distribution of ions present; median mobility k(med).

($n = F(1/k_*)$ or $n = \phi(k_*)$) could be plotted, and from these the distribution functions $f(k) = dN/dk$ were computed. These functions are shown in Figs. 2 and 3. As it was not justifiable to indicate any finer structure of the distribution function, $f(k)$ is shown as a histogram. Even if it were possible to discern some fine structure in the characteristics, and if this fine structure could also be accounted for, the method used cannot be regarded in its present state capable of such fineness. The characteristic general properties of the distribution function are not affected by this, and these properties are the following:

VOL. 4 1953

(1) The occupied range of mobilities is comparatively wide. It begins at c. 0·4–0·5 cm sec^{-1}/volt cm^{-1} and reaches c. 2 for positive ions. The highest mobility limit of negative ions is usually c. 2·5, in some cases somewhat more. This fact is not shown in the figures, but the columns for the higher mobilities are adequately heightened in these cases.

(2) The distribution function may be divided into two components: ions with lower mobilities are included in one, ions with higher mobilities in the other. In some cases a structure may be assigned to these components. The component for the lower mobilities is commonly higher than that for the higher mobilities. For positive ions the latter is wider. To avoid unnecessary complications, the group of negative ions with mobilities below 0·5, which occurred after sunrise, is not shown in the figures.

Median mobilities

In the course of the computation of the distribution function it is easy to obtain a mobility which may be called the *median mobility*, using a definition of statistics. This mobility is defined by the equations

$$k_1 < k\ (\text{median}) < k_2,$$

$$\int_{k_1}^{k(\text{med})} f(k)\,\mathrm{d}k = \int_{k(\text{med})}^{k_2} f(k)\,\mathrm{d}k = \tfrac{1}{2}\int_{k_1}^{k_2} f(k)\,\mathrm{d}k \equiv \tfrac{1}{2}N\Big]_{k_1}^{k_2},$$

where k_1 and k_2 are the limit mobilities of the distribution function. The median mobility $k(\text{med})$ divides the area under the distribution curve into two equal fractions. The values obtained, shown in Figs. 2 and 3, are in good agreement with the mean values of the mobility of atmospheric small ions given in the literature. Our average $k^{(+)}_{\text{med}} = 1{\cdot}01$, $k^{(-)}_{\text{med}} = 1{\cdot}13$; the mean mobilities of atmospheric small ions on the continent are $k^{(+)} = 0{\cdot}96$, $k^{(-)} = 1{\cdot}05$, according to the new edition of *Landolt-Börnstein* (Israël, 1952).

On August 7–8, 1952

The *weather conditions* were not very favourable for the measurements. On August 7 a region of low pressure situated over central Norrland was displaced, together with the associated rain fall, to the NE. Another region with rain over the Netherlands and North Germany moved northward to South Sweden. (On August 8 a cold NW air-stream with showers of rain appeared over N. Norrland, being especially marked after midday. The local weather conditions were as follows.) At 19 h there was a light NE breeze, the horizon was clouded in the W, and the sun was behind the clouds. The sunset was clear; at 20 h 20 m there was a light NE breeze, with large cumulus in the W; at 21 h 45 m a halo appeared round the moon. On August 8 at 2 h 45 m there was a light breeze from the E, mist in the Fyris valley, a light halo round the moon; at 4 h 05 m the mist from the Fyris ascended to higher layers, the breeze was slight N, NNE, the sunrise was obscured by the mist; at 4 h 12 m the mist became thicker near the ground. At 5 h there was dense mist all around; at 5 h 10 m the density of the mist decreased, and the sun shone through. At 5 h 40 m there was only light mist; at

5 h 47 m none. At 8 h 28 m there was slight stratus and the cloudiness gradually increased. Finally, the sky was clouded.

The temperature was recorded with a thermograph at ground level, and the relative humidity with a hair hygrograph. As shown in Figs. 4 and 5, there is a

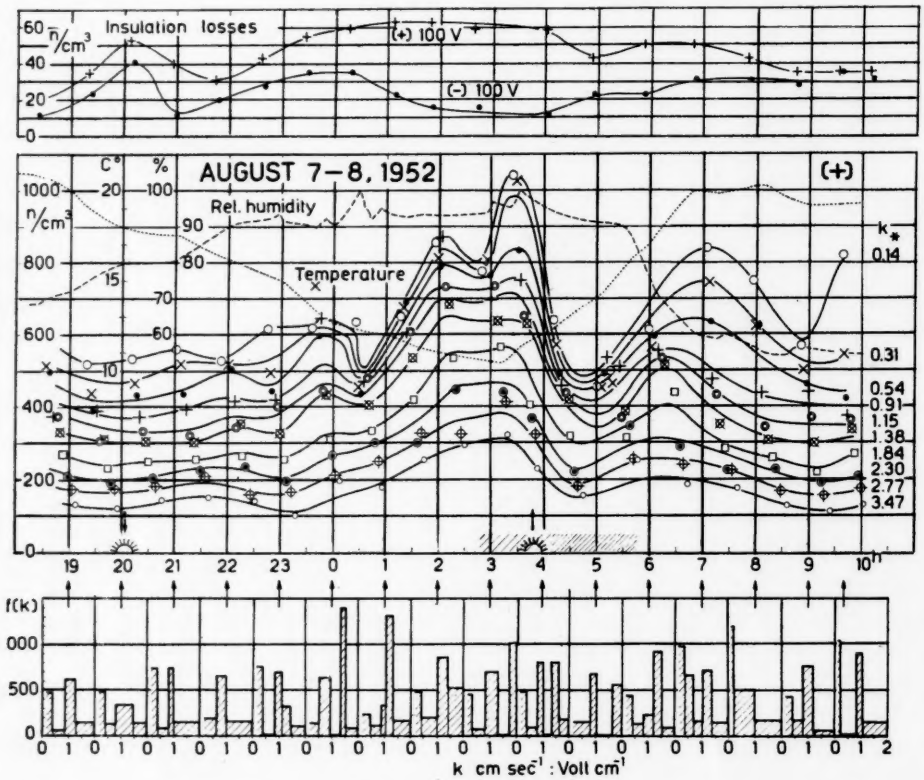


Fig. 4. Insulation losses; variation of the number of *positive* ions recorded at chosen limit mobilities k_* of an aspiration condenser; $f(k)$ mobility distribution of ions present.

perceptible correlation between the concentration of ions and the temperature and the relative humidity. The correlation with the relative humidity seems to be closer.

The curves of number of ions/time

In order to obtain more evidence on the structure of the mobility bands established by the earlier measurements, additional adjusted limit mobilities k_* were introduced between those used earlier. Altogether, 10 adjusted k_* were used during this series of measurements. It turned out on calculating, however, that the highest k_* were not usable in some cases, because of uncertainty in the numbers recorded, which were comparatively small.

A difference concerning the saturation state may be seen when the curves

VOL. 4 1953

obtained on July 7–8 and on August 7–8 are compared. Saturation is not reached on August 7–8, this being especially marked during the later morning hours. Ions with lower mobilities must have been present in greater concentration during this night. In spite of some deviations at certain points, it seems that the curves drawn through the points give an adequate description of the state. The mentioned deviations may be explained by considering more violent variations, such as have been observed in other series of measurements, especially during disturbed

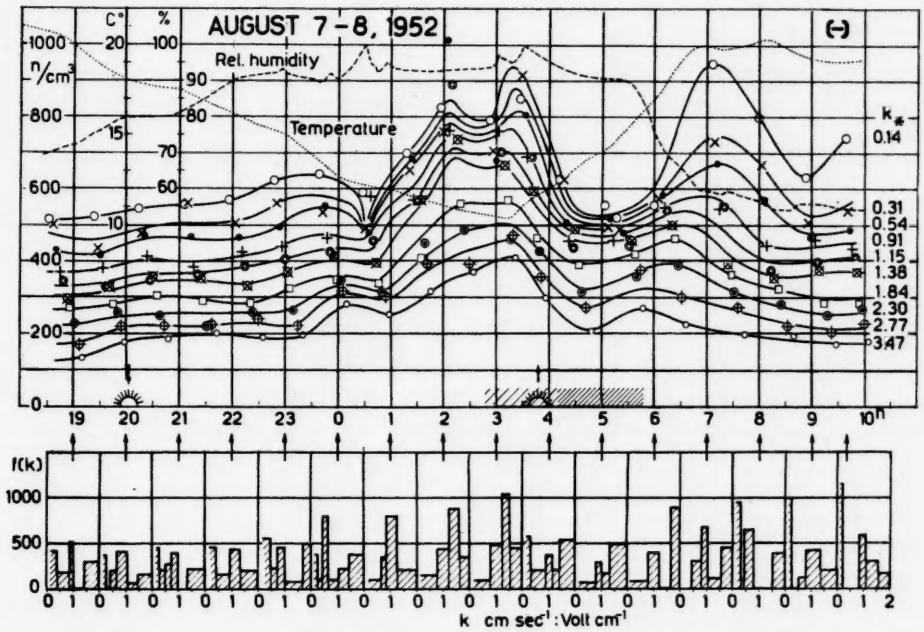


Fig. 5. Variation of the number of *negative* ions recorded at chosen limit mobilities k_* of an aspiration condenser; $f(k)$ mobility distribution of ions present.

weather conditions. The fact that the conditions were not tranquil in this case—though not so much that it would not be possible to discern the shorter time variations—may be seen from the trends of the temperature and the relative humidity. During the night on August 7–8, the amplitude of the temperature curve was less than half of that on July 7–8. The concentration of ions was also half as great—which fact also indicates disturbances.

Mobility distribution of the ions

Because of the greater number of k_* chosen in the same interval of mobilities, the structure of $f(k)$ was finer. But because of the troubled conditions, which caused more violent variations and a lower level of the number of ions present, the time curves of the distribution function were not so definite as in the previous case. Despite these drawbacks, the following properties of the distribution function may be discerned:

(1) The occupied range of mobilities is nearly the same, from 0·4, 0·5 to 2 for positive ions and to 2·5 for negative ions, but

(2) besides this range, there were also ions in a range of lower mobilities, and these two ranges were separated from each other.

(3) In the region of mobilities from 0·4, 0·5 to 2, 2·5, the distribution function may be divided into two parts which were completely separated in some cases.

4. DISCUSSION

The measurements considered above seem to show that it is possible with very simple means to measure the mobility of the small ions in the atmospheric air under favourable conditions. The computed mobilities were not constant, but varied with time between certain limits. In addition, conclusions could be drawn that the ions present were distributed over a range of mobilities. How may these facts be brought into accordance with commonly used conceptions? It seems that a confusion exists concerning the atmospheric ions, owing to the application of ideas not appropriate to them. The ideas and definitions must therefore be clarified to begin with.

Definitions and facts

The classification of the atmospheric ions is founded upon their mobilities, the only quantity which can be measured with reliability. Usually we consider a gaseous ion to be an atom or a molecule with one or more electrons missing from the outer shell or shells (positive ion), or with one or more additional electrons in the outer shell or shells (negative ion). Such ions, which might be called *atomic ions* or *molecular ions*, may be observed in pure gases at lower pressures, and they are well investigated. But if these ions enter or are produced in gases at higher pressures (atmospheric pressure, for instance), their life is only a fraction of a second, and their mobilities are reduced to 1–2 cm sec^{-1}/volt cm^{-1}. One of the explanations of this reduction in the mobility is based upon the assumption that other molecules become associated with the simple monomolecular ion when it enters in a gas atmosphere at higher pressure, thus forming a *cluster-ion* (ZELENY, 1930). Under conditions of high purity the mobility of a cluster-ion is comparatively well defined (for example $k_+ = 2{\cdot}0$, $k_- = 2{\cdot}45$ in air at atmospheric pressure (SCHILLING, 1927). The cluster-ion may therefore be assumed to be a comparatively stable aggregate. But now if the conditions are not of the highest purity—very small traces of impurity are sufficient—the mobility of the cluster-ion is reduced once again, in fact down to a region of mobilities with an average value of 1 or a little more. The average values of the mobility of natural small ions in the free atmospheric air do not differ much from 1 (HESS, 1926; ISRAËL, 1952). How can these facts be reconciled? It seems quite clear that atmospheric small-ions are not monomolecular ions, but it also seems that they are not pure cluster-ions. There exist two views about the mobility of cluster-ions (PRZIBRAM, 1933). The mobility of a cluster-ion is discrete and distinct according to the first view. According to the other, the cluster-ions have mobilities statistically distributed over a certain range of mobilities, between certain limits which are dependent on the impurities. Not all the methods of measuring the mobilities are

VOL. 4 1953

suitable for determination of the distribution mentioned, and moreover, some of the measuring methods have the disadvantage that they introduce a fictive dispersion of the mobilities over a larger range because of diffusion. This is not the place for a critical consideration of such questions, and that must be left to another occasion.

A compromise may be made between the two views in question. The authors who promulgate the existence of discrete mobilities for the small ions have emphasized the conditions of purity in their experiments, and it is very possible that under such conditions the mobilities are discrete or, in the event of a statistical distribution, that the limits of the distribution function are very narrow. On the other hand, it is not feasible that the extension of the measured mobility range established by other authors is due only to the measuring devices, as some of these authors have even established a separation of the distribution function into two parts (ERIKSON, 1929; TYNDALL *et al.*, 1928; ZELENY, 1930, 1931). It may be questioned whether the conditions were the same in all cases. Maybe we can assume that pure cluster-ions were involved when the experiments were carried out under conditions of purity, and that products of subsequent transformations caused by the impurities were included in the measurements when the conditions concerning purity were not so rigorous. It seems that the transformed small ions also form a definite group, with specific properties, though they have not been considered worthy of investigation until now. It is first necessary to assign a name to this class of ions. They are evidently formed from the initial small-ions. It might be supposed that they may perhaps also be transformed continuously into less mobile intermediate and large ions. It seems that is not so. Evidence for this conclusion is contained in the results obtained by the authors quoted, who have established the second lower mobility. Our results also show that the range of mobilities between 2, 2·5–0·4, 0·5 is delimited from the region of the following intermediate ions. Similar well-defined groups of free atmospheric ions with mobilities between 1·62 and 0·9 cm sec^{-1}/volt cm^{-1} and in the range 0·9 to 0·4, distinct from the following intermediate ions, were established by YUNKER (1940), who assumed that the ions in the range 0·9–0·4 were the same as found by CHAPMAN (1937) in spraying and bubbling of liquids. Furthermore, the median mobilities computed by us, average $k^{(+)}_{med} = 1{\cdot}01$, $k^{(-)}_{med} = 1{\cdot}13$, agree very well with the average mobilities of atmospheric small ions given by other authors (HESS, 1926; ISRAËL, 1952); and it is evident that commonly used measuring methods would not be able to establish a distribution of mobilities among the small ions. Such are the reasons for including the ions in question in the small-ion group, but in order to avoid confusion with the initial small ions, the transformed ions might be called *affected small ions*, or *affected cluster ions*, because the "pure" small ions are mainly affected by impurities. The question of how the initial small ions may be transformed into affected small ions must be left unanswered for the time being. It has been shown (ERIKSON, 1929; TYNDALL, 1928) that in the air the ionizing process itself gives rise to molecules which transform the initial ions. Our first intention was to clarify the confusion existing up to now concerning the small ions in the atmospheric air. This led to a clearer formulation of the problems associated with this matter, and these problems have been found worthy of investigation.

Application to our results

Figs. 2 and 3 show columns which may be attributed to initial small-ions and to affected small-ions. To begin with, when emanation cannot be expected to accumulate in the air, the band of mobilities spreads from 0·5 to 2 (the first three columns for negative ions and the first for positive ions), or the columns with mobilities from 0·5 to 1 are heightened (the second and third columns for positive ions). Initial small ions were formed, but not so many, and they were transformed into affected small ions. After midnight the height of both the columns is higher, and the highest mobility of the group of affected ions is displaced to higher values. More new initial ions had been formed, and they could not be transformed so rapidly into affected small ions. At the end of the series the importance of affected small ions once more increases anew. The considerable amount of ions with higher mobilities after sunrise may be explained by the fact that at this time an increase in the concentration of large ions may be expected, as shown during our earlier measurements and under similar weather conditions (NORINDER and SIKSNA, 1952*a*), and the content of radio-active substances in the air may increase with increase in the concentration of nuclei, as shown by BURKE and NOLAN (1950).

A similar trend may be seen in Figs. 4 and 5, only the groups for initial and affected small ions were often completely separated here, because of the higher resolving power in this case. At lower mobilities, intermediate ions also occurred here. Deviation from the general trend may be explained by the weather conditions which disturbed the measurements, but the general shape remains. Indications that small ions may be affected in this way have also been observed for ions formed in corona discharges (SIKSNA, 1953).

Acknowledgments—We thank the Swedish National Science Research Council for placing means at the disposal of the Institute to perform this investigation. The authors gratefully acknowledge the assistance of Mr. A. METNIEKS and Dr. H. A. WATSON in the observations and calculations.

REFERENCES

AGARWALA, K. S. (1951) The mobility of the small ions of the atmosphere at Poona. Indian J. Met. Geophys. **2**, 63–65

BURKE, T. and NOLAN, J. J. (1950) Observations on the Radium A content of the atmosphere. Proc. Roy. Irish Acad. **53** (A), No. 9, 145–161

CHAPMAN, S. (1937) Carrier mobility spectra of spray electrified liquids. Phys. Rev. **52**, 184–190

ERIKSON, H. A. (1929) Factors affecting the nature of ions in air. Phys. Rev. **34**, 635–643

FEDOROW, E. K. (1952) Novy metod izsledovanija ionnovo spektra w atmosfere. Doklady AN SSSR **82**, 717–718

GERDIEN, H. (1903) Die absolute Messung der elektrischen Leitfähigkeit und der spezifischen Ionengeschwindigkeit in der Atmosphäre. Phys. Z. **4**, 632–635

HESS, V. (1926) Die elektrische Leitfähigkeit der Atmosphäre und ihre Ursachen. Braunschweig: Fr. Vieweg & Sohn A. G.

ISRAËL, H. (1952) Luftelektrizität. Landolt-Börnstein, Zahlenwerte und Funktionen, 6. Aufl., Bd **3**, Berlin-Göttingen, Heidelberg: Springer-Verlag, p. 711.

MACHE, H. (1903) Eine einfache Methode, die Geschwindigkeit von Gasionen zu bestimmen, welche weder hohes Molisierungs-noch Regenerierungsvermögen besitzen. Phys. Z. **4**, 717–721

NORINDER, H. and SIKSNA, R. (1950) Variations in the density of small-ions caused by the accumulation of emanation exhaled from the soil. Tellus **2**, 168–172

VOL.
4
195

NORINDER, H. and SIKSNA, R. (1952*a*) Variations of the concentration of ions at different height near the ground during quiet summer nights at Uppsala. Kungl. Svenska Vetenskapsakademiens Ark. f. geofys. **1**, Nr 19, 519–541

NORINDER, H. and SIKSNA, R. (1952*b*) Ions produced in the air at atmospheric pressure by ultra-violet light from a quartz-mercury arc. Kungl. Svenska Vetenskapsakademiens Ark. f. fys. **5**, Nr 23, 471–491

O'DONNEL, G. A. (1952) Electric conductivity and small ion concentration of the atmosphere at one metre above ground and conductivity at ground level. J. Atmosph. Terr. Phys. **2**, 201–215

PRZIBRAM, K. (1933) Die Ionen in Gasen. Handb. d. Phys. herausg. von H. GEIGER u. K. SCHEEL, 2. Aufl. **22, 1,** 343–423

SCHILLING, H. (1927) Ueber die kleinsten Elektrizitätsträger in Gasen. Ann. Phys. (Leipzig) (4) **83**, 23–79

SIKSNA, R. (1952*a*) Positive ions formed by an open electric heater. Kungl. Svenska Vetenskapsakademiens Ark. f. fys. **5**, Nr 25, 531–543

SIKSNA, R. (1952*b*) Mobility spectra of ions formed in a room by negative corona discharge. Kungl. Svenska Vetenskapsakademiens Ark. f. fys. **5**, Nr 26, 545–564

SIKSNA, R. (1953) Mobility spectra of ions formed by positive corona discharge. Kungl. Svenska Vetenskapsakademiens Ark. f. fys. **6**, Nr 28, 279–286

TORRESON, O. W., PARKINSON, W. C., GISH, O. H., and WAIT, G. R. (1949) Ocean Atmospheric-Electric Results. Washington DC: Carnegie Institution of Washington Publication

TYNDALL, A. M., GRINDLEY, G. C., and SHEPPARD, P. A. (1928) The mobility of ions in air. Part V. The transformation of positive ions at short ages. Proc. Roy. Soc. London (A) **121**, 185–194

VASSAILS, G. (1949) Recherches sur les gros ions dans les gas. Ann. Phys. (Paris) (12) **4**, 125–217

WEGER, N. (1935) Eine Verbesserung der Methodik der Kleinionen-Zählung. Phys. Z. **36**, 15–20

YUNKER, E. A. (1940) The mobility-spectrum of atmospheric ions. Terr. Magn. **45**, 127–132

ZELENY, J. (1930) The mobilities of ions in dry and moist air. Phys. Rev. **36**, 35–43

ZELENY, J. (1931) The aging of ions in air and nitrogen. Phys. Rev. **38**, 969–976

OL.
4
953

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 106 to 111. Pergamon Press Ltd., London

Mobility of small atmospheric ions in the air from the ground at Uppsala

R. SIKSNA

Institute of High Tension Research, University of Uppsala, Sweden

(*Received* 4 *July* 1953)

ABSTRACT

By counting the ions in the air below a sheet covered over the ground, using a simple aspiration condenser, it is shown that the small ions must be divided with respect to their mobilities into two fractions, which may be attributed to the initial small ions and the affected small ions. The ions counted were formed by the radioactive substances exhaled from the soil. It was observed that the mobility distribution of the ions depended on the speed with which the air was sucked through the intake pipe to the counter.

1. INTRODUCTION

In the layers of atmosphere near the ground the small ions are mainly produced by radioactive substances. The content of the radioactive substances is higher than normal in the surroundings of this Institute (NORINDER *et al.*, 1952). An increase has also been observed in the concentration of small ions caused by accumulation of emanation exhaled from the soil during quiet summer nights (NORINDER and SIKSNA, 1952). During measurements of the mobility of ions present in the open air under such conditions (NORINDER and SIKSNA, 1953), the idea arose that it might be possible to observe an enhanced concentration of ions in the air near the ground by covering the ground and so preventing the exhaled emanation from dispersing in the higher layers of the atmosphere. And in fact, when the ground was covered with an impermeable screen the ion concentration was considerably higher in the air sucked from the space between the ground and the sheet than it was in the open air. The mobilities of these ions were investigated, in accordance with the view indicated in (NORINDER and SIKSNA, in press), that the ions commonly called small atmospheric ions consist of initial small ions and affected small ions.

VOL. 4 1953

2. MEASURING ARRANGEMENT

The place for measuring was that used earlier when measuring atmospheric ions at this Institute (NORINDER and SIKSNA, 1952; NORINDER and SIKSNA, 1953). The measuring devices were placed in an observation hut (Fig. 1). The measuring apparatus was the same as used for determining the mobility of small ions in the open air during summer nights (NORINDER and SIKSNA, 1953): an ion-counter *W* with the Weger aspiration condenser. This counter was connected with the suction box *B*. An Ebert counter *E* was connected with the suction box in order to maintain a suitable air-flow through the box. The air was sucked into the suction box through a 3 m long pipe *P*, which consisted partly of a brass tube of 3 cm diameter and partly of a flexible metal tube wrapped with insulating tape to prevent leakage. The suction pipe was partly buried in the earth, so that its open end came out a little above the ground, as shown in Fig. 1. Over the open end of

the pipe was placed a rectangular sheet of zinc 175 × 270 cm, and with the edges bent down so that a 10 cm high closed space was built close to the grass-covered ground. The ions counted came from this space. In order to determine the losses of ions during passage of air through the pipe, the latter could be bent so that its open end came under the protecting box on the roof of the hut, as shown in Fig. 1. By closing and opening the slides *Sp* and *Su* it was possible to suck the air supplying ions for counting from the open air at the level of the roof alternately through the pipe *P* and through the wide opening *T*. For determining the mobility of the ions present, the method of periodical alternation of chosen limiting mobilities of the ion-counter was used (NORINDER and SIKSNA, 1953).

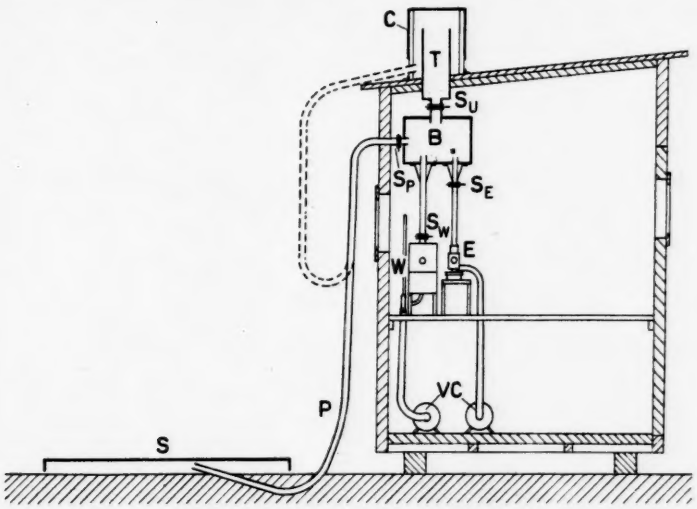


Fig. 1. The apparatus used: *W* small ion counter with a Weger condenser; *E* Ebert ion counter; *B* suction box; *P* suction pipe; *S* sheet covering the ground; *T* opening to the open air; *C* protection box; S_u, S_p, S_u, S_E slides; *VC* vacuum cleaners.

3. RESULTS

The concentration of ions in the air sucked from the space below the zinc screen was considerably higher than that in the open air. With the given measuring arrangement it was highest, c. 10 times higher than in the open air, when air was slowly sucked into the suction box (Φ = 400 cm³/sec) only through the measuring counter *W*. When the air was sucked into the suction box with the Ebert counter as well, the concentration was lower, but even in this case it was 2–3 times greater than in the open air, even if the air-flow through the suction box was the maximum used—5400 cm³/sec. There was no indication that the air with the greater concentration of ions might be exhausted from the space below the sheet, so that the concentration might decrease with time. Although variations of concentration were observed, as shown in Figs. 3 and 4, this was not due to exhaustion of the supply of ions. These variations with time were not investigated in detail, because if they are to be determined more precisely, the screen must also be made so that the air exhaled from the ground is not diluted with other air not containing so

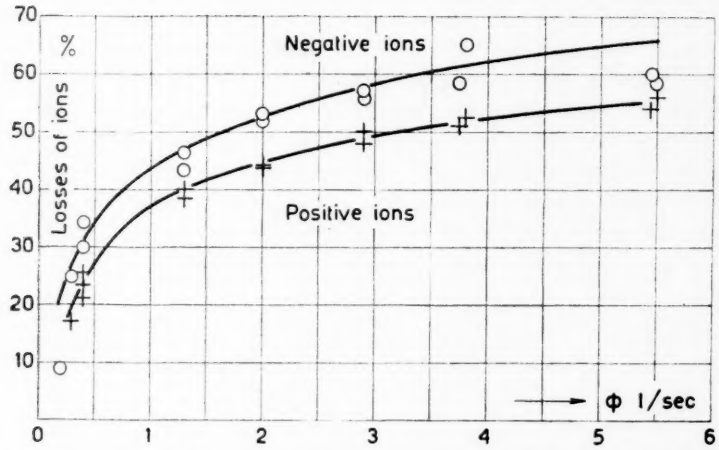


Fig. 2. Losses of concentration of ions due to passage of the air through a pipe 30 mm diameter and 3 m long at different suction speeds Φ.

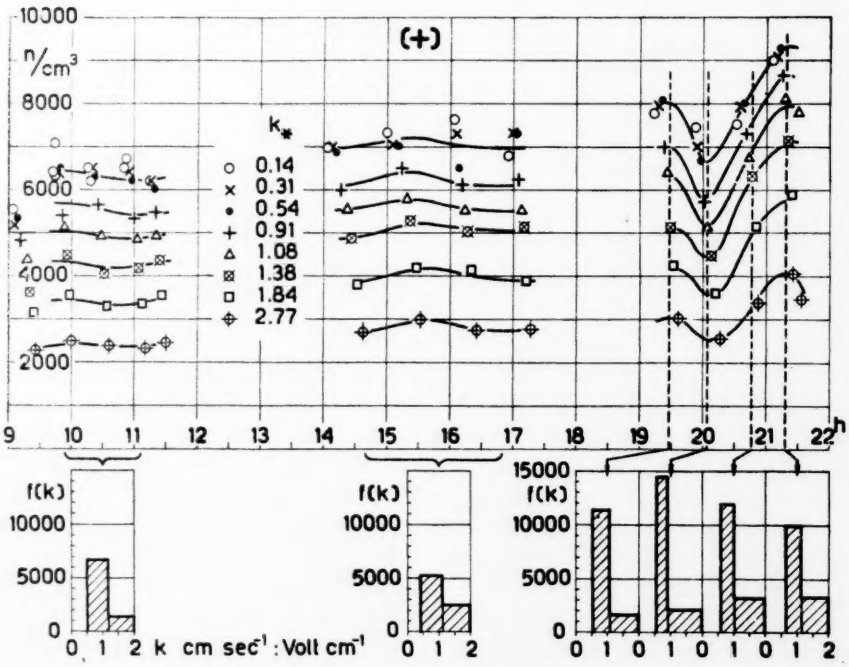


Fig. 3. The number of positive ions in the air from the covered ground at chosen limiting mobilities k_* of the counter and with an air-flow of 400 cm³/sec through the suction box. The distribution function $f(k)$ of the ions with respect to the mobilities.

VOL. 4 1953

many ions. Reduction of the concentration of ions when the air-flow is more rapid may be explained by reference to the loss of ions during passage of air through the pipe. The variation of this loss with the speed of air-flow through the pipe is shown in Fig. 2, which gives results of measurements when the open end of the suction pipe was placed on the roof, and the concentration of ions in the open air was measured by passing the air alternately through the pipe P and the opening T. From approx. $\Phi = 0{\cdot}8$ 1/sec the air-flow in the pipe is turbulent, and this fact may influence the loss of ions. As shown in Fig. 2, the loss of negative ions is greater than that of positive ions, which may be explained by the higher mobility of negative ions. Because of the comparatively low concentration of ions in the

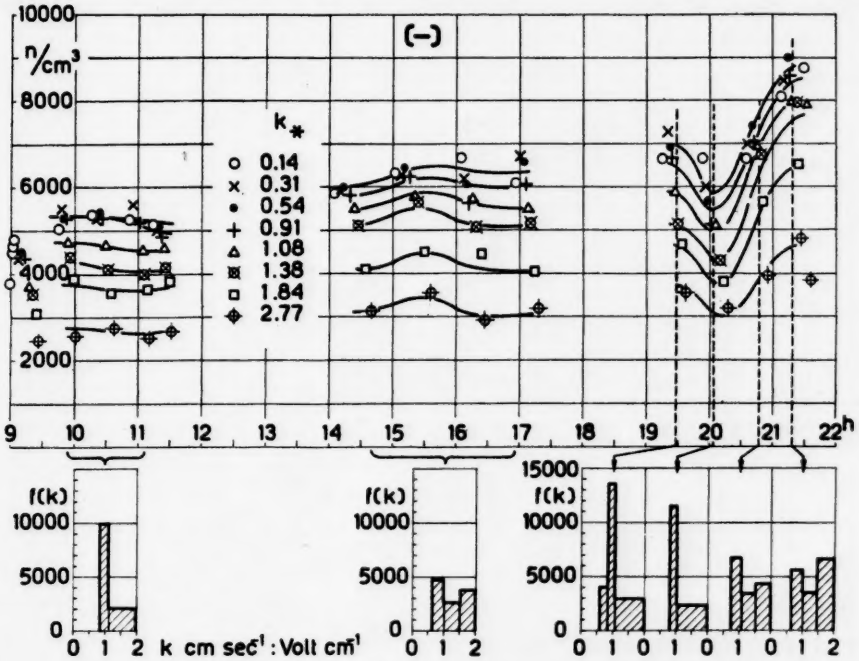


Fig. 4. The number of negative ions in the air from the covered ground at chosen limiting mobilities k_* of the counter and with an air-flow of 400 cm³/sec through the suction box. The distribution function $f(k)$ of the ions with respect to the mobilities.

open air, there was a large random error in the measurements of ion losses with the arrangement mentioned, especially at lower suction rates. In Figs. 3 and 4 the curves of the number of ions measured at the chosen limiting mobilities k_* of the counter condenser are shown for the air sucked from the space below the sheet at $\Phi = 400$ cm³/sec. Though a random variation may be seen, the numbers of ions counted at the three lower k_* values—0·14, 0·31, 0·54 cm sec^{-1}/volt cm^{-1}—could be assumed to be upon one curve, and may therefore be taken to represent the saturation at these limiting mobilities. For the first two measuring periods, averages of the measured numbers of ions were found for each limiting mobility k_*, and with the aid of the $n = F(1/k_*)$ and $n = \phi(k_*)$ characteristics the distribution function $f(k)$ with respect to the mobilities was computed. For the third

measuring period, the greater variations in the figures made it necessary to compute the mobility distribution of ions for four particular instants from the intersection of vertical straight lines with the curves giving the number of ions counted. As may be seen in the figures all the $f(k)$ obtained are similar to that obtained for ions in the open air (NORINDER and SIKSNA, 1953). Furthermore, as in that case, no attempt has been made to distinguish a finer structure of the distribution function.

The mobility distribution of the present ions is as follows: (1) The occupied range of mobilities is comparatively wide. It begins at c. 0·4–0·5 cm sec^{-1}/volt cm^{-1} for positive ions, and at c. 0·6–0·8 cm sec^{-1}/volt cm^{-1} for negative ions, and reaches c. 2 cm sec^{-1}/volt cm^{-1} for positive ions. The upper limit of mobility for negative ions is usually c. 2·3–2·5 cm sec^{-1}/volt cm^{-1}. This fact is not shown in

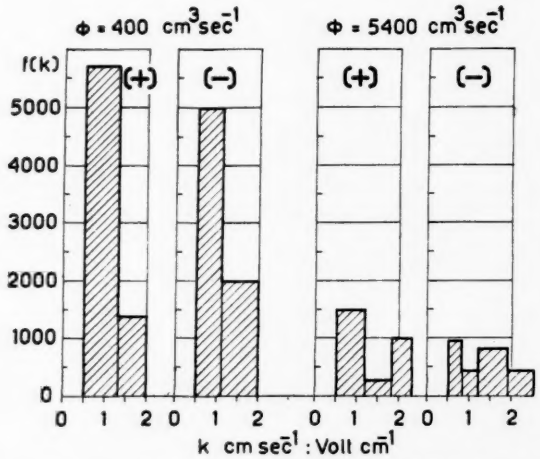


Fig. 5. The distribution function $f(k)$ of the ions with respect to the mobilities at the suction speeds $\Phi = 400$ cm³/sec and $\Phi = 5400$ cm³/sec through the suction box.

the figures, but the columns for higher mobilities are adequately heightened in these cases. (2) The distribution function may be divided into two components: ions with lower mobilities are included in one, ions with higher mobilities in the other. In some cases, a structure may be detected in these two components. The component representing lower mobilities is higher than that representing the higher. For positive ions the component representing lower mobilities is wider. At the ends of the observation periods considered, the component for higher mobilities increases—especially in the case of negative ions, where a higher column may be observed. Similar distribution functions were obtained from other series of measurements, by using the same air-flow $\Phi = 400$ cm³/sec and another of $\Phi = 5400$ cm³/sec through the suction box, as shown in Fig. 5. When the air-flow was greater, the number of ions was smaller, because of higher losses caused by the turbulent air-flow in the pipe; but the component representing higher mobilities was raised in this case for both the negative and the positive ions.

It will be seen that the observed facts confirm the idea (NORINDER and SIKSNA, 1953) that in the region of mobilities attributed to small atmospheric ions two

VOL. 4 1953

kinds of small ions must be accepted: the *initial small ions* and the *affected small ions*. The proportion of initial small ions may increase if the total concentration of ions increases because of the formation of new ions, as shown in Figs. 3 and 4 during the last period when the total concentration increased. The proportion of initial small ions may also be greater if the air from the space between the ground and the sheet is transported to the counter with greater velocity ($\Phi = 5400$ cm^3/sec, Fig. 5), because in this case the ions formed by the radioactive substances in the air conveyed through the pipe will enter the counter a shorter time after their formation. In order to obtain greater clarity on the problems dealt with, particularly concerning the shape of the distribution function, a more precise method of measuring ionic mobilities must be used.

Acknowledgments—The author is grateful to the Director of the Institute of High Tension Research of the University of Uppsala, Professor H. NORINDER, for providing him with facilities to carry out this work and for his interest in its progress.

Thanks are also due to the Swedish Natural Science Research Council for placing means at the disposal of the Institute to perform this investigation. The author gratefully acknowledges the assistance of Mr. A. METNIEKS and Dr. H. A. WATSON in the observations and calculations.

REFERENCES

NORINDER, H., METNIEKS, A. and SIKSNA, R. 1952 Radon content of the air in the soil at Uppsala. Kungl. Svenska Vetenskapsakademiens Ark. f. geofys. **1**, Nr 21, 571–579

NORINDER, H. and SIKSNA, R. 1952 Variations of the concentration of ions at different heights near the ground during quiet summer nights at Uppsala. Kungl. Svenska Vetenskapsakademiens Ark. f. geofys. **1**, Nr 19, 519–541

NORINDER, H. and SIKSNA, R. 1953 Mobility of atmospheric small ions during summer nights at Uppsala. J. Atmosph. Terr. Phys. **4**, 93

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 112 to 123. Pergamon Press Ltd., London

The altitudes of the luminous layers in the earth's atmosphere.

D. R. Bates and A. Dalgarno

Department of Applied Mathematics, Queen's University, Belfast

(*Received* 6 *August* 1953)

ABSTRACT

The claim of the observers that the main features of the airglow (other than the Meinel hydroxyl bands) originate from altitudes as great as several hundred kilometres is disputed. Consideration is given to the various excitation mechanisms which might conceivably be operative but even on the most favourable assumptions none appears likely to be sufficiently effective at such altitudes. In the case of certain of the emissions, such as the *D*-lines of sodium, it is contended that the theoretical arguments are unequivocal.

1. Introduction

Much effort has been devoted to the determination of the location of the layers responsible for the airglow. The results are far from being in harmony, but most observers currently advocate that a number of the main spectral features originate from an altitude of several hundred kilometres. If correct, this is of the utmost importance in connection with the theory of the excitation processes, and indeed with upper atmospheric physics in general; but it is not a conclusion which should be accepted lightly. After briefly summarizing the published results we shall advance arguments against the altitudes being as great as claimed.

VOL.
4
195

2. Published Results

Measurements have been carried out with spectrometers and with photometers. There has been a tendency of late to abandon spectrometers since it is now considered that the correction for the background continuum is too uncertain. We shall therefore quote only a few of the results obtained with this type of instrument and concentrate our attention on the investigations carried out with photometers, the development of the use of which is principally due to Abadie, Vassy and Vassy and to Roach and his associates. Table 1 lists the emissions studied and the altitudes found by various observers together with those recommended by Barbier (1952) in a recent critical review of the work which has been done.

The quantity that is actually measured is $I(\theta)/I(0)$ the ratio of the intensity at zenith angle θ to the intensity at zenith angle zero. Knowing the value of this ratio for any given spectral feature it is possible, in principle, to determine the distance above the earth's surface of the luminous layer concerned by the application of the classical formula of van Rhijn (1921). Corrections for atmospheric extinction and for extra-terrestrial radiation have however to be introduced; and as these are subject to some uncertainty it is natural to enquire if errors in them could account for the deduced altitudes being so much greater than the expected altitudes. To illustrate the position Table 2 was compiled using formulae due to Barbier (1944). It applies to a layer at the 100 km level (which is in the general region favoured by most theorists) and gives the calculated ratios $I(75°)/I(0)$ and

Table 1. *Summary of observational results*

Emission	*Reference*	*Instrument* used*	*Altitude (kms)*	
			Value derived in original paper cited	BARBIER'S *estimate from survey of all observations*
λ5893 of NaI	DUFAY and TCHENG (1947b)	*S*	80	Approximately 300
	ROACH and BARBIER (1950)	*P*	310	
	ROACH and PETTIT (1952)	*P*	275	
λ5577 of OI	DUFAY and TCHENG (1947b)	*S*	100	250(±30)
	ROACH and BARBIER (1950)	*P*	110	
	BARBIER, DUFAY and WILLIAMS (1951)	*P*	215	
	ROACH and PETTIT (1951)	*P*	250	
	ROACH and PETTIT (1952)	*P*	200	
	ABADIE, VASSY and VASSY (1949)	*P*	400 to 1000	
λ6300 of OI	DUFAY and TCHENG (1947b)	*S*	180	Much greater than 250, perhaps about 500
	ROACH and PETTIT (1952)	*P*	275	
	ABADIE, VASSY and VASSY (1949)	*P*	400 to 1000	
Herzberg bands of O_2	BARBIER (1947b)	*S*	350	200 to 250†
	BARBIER (1953)	*P*	100 to 400	
Vegard-Kaplan bands of N_2	BARBIER (1947b)	*S*	900	350(±100)
Meinel bands of OH	ROACH, PETTIT and WILLIAMS (1950)	*P*	70	Close to 70
	HURUHATA (1950)	*P*	300	
λ5180 continuum	BARBIER, DUFAY and WILLIAMS (1951)	*P*	460	460(±50)

* *S* denotes spectrometer and *P* photometer.

† In his review BARBIER (1952) gave the probable altitude for the Herzberg band system to be 100 km. This is not outside the range allowed by his latest observations (1953), but he now favours an altitude of 200–250 km because he considers that the variation in intensity of the system is so similar to that of λ5577 that the two luminous layers must lie close together. It may be remarked that correlations between the variations of λ5577 and the λ5180 continuum (BARBIER, DUFAY and WILLIAMS, 1951), and between the variations of λ6300 and λ5893 (DUFAY and TCHENG, 1947a) have also been reported. Though undue weight should not be attached to them, these correlations should be borne in mind in the subsequent discussion.

$I(80°)/I(0)$ for various assumed values of τ, the atmospheric extinction coefficient,‡ and of ρ, the zenith intensity due to extra-terrestrial sources over the zenith intensity due to the layer. Consider now some representative observational data. BARBIER, DUFAY and WILLIAMS (1951) find that $I(75°)/I(0)$ is 2·20 for λ5577; and ROACH and BARBIER (1950) find that $I(80°)/I(0)$ is 2·21 for λ5893.§ If haze is

‡ For simplicity τ has been taken to include both Rayleigh scattering and true absorption though strictly the two effects should be treated separately. The errors introduced by the approximation are insignificant.

§ It should be noted that the measured ratios show a wide scatter: for example, in a survey of the sky which ROACH and BARBIER give in detail as typical $I(80°)/I(0)$ ranges between 2·09 and 3·28 for the different azimuths.

completely absent the value of τ should be about 0·11/atmosphere at the former wavelength, and about 0·09/atmosphere at the latter. It is judged too, by the authors responsible, that the results cited correspond to a ρ of zero. On comparison with the figures presented in Table 2 it will be seen that they can only be reconciled with an altitude of approximately 100 km by assuming that atmospheric extinction is grossly underestimated, that extra-terrestrial radiation is very incompletely eliminated, or that there is an unknown source of error; and none of these assumptions would seem at all plausible in view of the care with which the measurements have been conducted.* It must be acknowledged therefore that the observational evidence is strong. In spite of this we believe that the problem is far from closed.

Table 2. Calculated values of $I(75°)/I(0)$ and $I(80°)/I(0)$ *for a layer at the* 100 *km level*

ρ	0·0	0·1	0·2	0·3	0·0	0·1	0·2	0·3
τ (/ atmosphere)	$I(75°)/I(0)$				$I(80°)/I(0)$			
0·00	3·24	3·03	2·86	2·72	4·09	3·81	3·57	3·38
0·03	3·02	2·83	2·67	2·54	3·64	3·40	3·19	3·02
0·06	2·82	2·65	2·51	2·38	3·27	3·06	2·88	2·72
0·09	2·65	2·49	2·36	2·24	2·96	2·77	2·60	2·47
0·12	2·49	2·35	2·22	2·12	2·69	2·52	2·38	2·26
0·15	2·36	2·22	2·10	2·01	2·47	2·32	2·19	2·08
0·18	2·23	2·10	2·00	1·91	2·28	2·14	2·02	1·93
0·21	2·12	2·00	1·90	1·82	2·12	1·99	1·89	1·80

VOL. 4 1953

3. Theoretical considerations concerning the altitudes

The source of the airglow has not yet been established beyond dispute and it is conceivable that some mechanism has been completely overlooked by the theorists. Caution must therefore be exercised in making deductions about the location of the luminous layers. Nevertheless in the case of some of the emissions the objections to an altitude of several hundred kilometres seem almost conclusive; and if the observed altitudes of these emissions are, in fact, incorrect the observed altitudes of the others must be regarded as at least suspect.

For convenience we shall discuss the different spectral features separately.

λ5893 *of* NaI

Free atomic sodium is an extremely rare constituent of the upper atmosphere: for example from a search of high dispersion stellar spectrograms for signs of the *D*-lines in absorption Sanford (1950) concluded that at night there is at most some 1×10^{10} Na atoms/cm^2 column. Because of this rarity Bates and Nicolet (1950a) were able to prove by general arguments, which need not be repeated, that if (as seems probable) some chemical reaction gives rise to the excitation, then the

* However since this was written Chamberlain and Oliver (1953) have reported that there is evidence that in at least some spectral regions starlight is much more important than has hitherto been believed.

altitude is certainly very much less than 300 km (cf Table 1) and indeed can scarcely be much above 70 km.

It has been shown (BATES, 1947) that near the 300 km level

$$n(\mathrm{Na}^+) \gg n(\mathrm{Na}) \tag{1}$$

where the n's represent the concentrations of the particles indicated. Consequently a natural suggestion to make is that the excitation is provided by some ionic reaction. Radiative recombination,

$$\mathrm{Na}^+ + e \rightarrow \mathrm{Na}(3\,^2P) + h\nu, \tag{2}$$

obviously can be ignored; but the charge transfer process,

$$\mathrm{Na}^+ + \mathrm{O}^- \rightarrow \mathrm{Na}(3\,^2P) + \mathrm{O}, \tag{3}$$

has been advocated by SAHA (1952) and merits some consideration. The simplest procedure is to examine whether it is possible to meet the essential requirement, that the rate of removal of the neutral sodium atoms must be equal to the rate of their formation which, if (3) is responsible for the airglow, must be at least as great as 8×10^7/(cm^2 column sec), the λ5893 photon emission rate (ROACH and PETTIT, 1952). Loss by chemical association has already been shown to be utterly inadequate (BATES and NICOLET, 1950a). The only other possibility is loss by ionization. Processes not in some way peculiar to sodium can clearly be disregarded. Though

$$\mathrm{Na} + \mathrm{O}^+ \rightarrow \mathrm{Na}^+ + \mathrm{O}' \tag{4}$$

and

$$\mathrm{Na} + \mathrm{N}^+ \rightarrow \mathrm{Na}^+ + \mathrm{N}' \tag{5}$$

may occur, neither are in close energy resonance so that the rate coefficients must be low (cf MASSEY and BURHOP, 1952); and even if they were 10^{-10} cm^3/sec, if the total number of Na atoms in the region of the 300 km level were 10^9/cm^2 column, and if $n(\mathrm{O}^+)$ or $n(\mathrm{N}^+)$ were 10^6/cm^3 (all of which are gross over-estimates) the Na^+ yield would only be 10^5/(cm^2 column sec), which is much too small.

The occurrence of fast electrons in the F-layers is sometimes postulated but, as may readily be verified, there is no possibility of appreciable excitation by them since the rather large cross section associated with

$$\mathrm{Na}(3\,^2S) + e \rightarrow \mathrm{Na}(3\,^2P) + e \tag{6}$$

(cf MASSEY and BURHOP, 1952) fails by many orders of magnitude to compensate for the rarity of the free sodium.

Credence cannot be attached to the hypothesis that the radiation arises from incoming Na atoms* for even if the kinetic energy were as low as 100 e volts (which corresponds to a velocity about equal to that of the earth in its orbit) the Doppler shift would be approximately 0·5 Å whereas according to BRICARD and

* Because of photo-action Na^+ ions must actually predominate (cf KAHN, 1950) but they need not be considered in the present connection as they would be directed towards the Poles. It may be remarked that BARBIER and PETTIT (1952) find some evidence that the intensity of the D-lines decreases with increasing distance from the Equator, and the first-named author (1947a) reports that they are not enhanced during low latitude aurorae.

KASTLER (1944) the shift of part of the actual airglow emission is some 0·01 Å or less. Moreover the incident stream would have to be absurdly dense to provide the observed intensity. Only a small fraction of the kinetic energy would appear in the *D*-lines (cf HERLOFSON, 1947), and even if the fraction were 0·1 (which is improbable) the Na atom concentration in the stream would have to be some 6/cm^3. Such a concentration is of course impossible to accept.

The case against λ5893 being emitted from an altitude even approaching 300 km would seem to be apodictic.

λ5577 and λ6300 of OI

The high abundance of atomic oxygen makes hazardous any attempt at enumerating all conceivable excitation processes. Because of this it would be improper to claim finality for any purely theoretical analysis of the situation; but the doubts that may be cast on the altitude given in Table 1 for the green line, and (to a lesser extent) on that given for the red line, are so grave that great weight must be attached to them.

CHAPMAN'S proposal (1931) that the excitation occurs during three-body association,

$$O + O + O \rightarrow O_2 + O(^1S \text{ or } ^1D), \qquad (7)$$

has some unattractive features but is widely accepted as the most plausible yet advanced. Any emission ensuing would be expected to be mainly from the region near the 100 km level since this is where n(O) is greatest. Admittedly the luminosity might lie at a higher level in the unlikely event of the rate coefficient, α, being a rapidly increasing function of the temperature. Account must however be taken of the absolute intensity. If the altitude were 250 km, as claimed, the local concentration n(O) could only be about 2×10^{10}/cm^3 even if the temperature were as high as 2000°K (cf BATES, 1951a); so that if the effective thickness of the active layer were 100 km, and α were 10^{-32} cm^6/sec (an improbably large value), the photon yield would be about 10^6/(cm^2 column sec), which is some hundred times smaller than the observed. The discrepancy would be even greater if the atmosphere should prove to be as cool and tenuous as measurements made during

Table 3. Results of rocket measurements

Altitude (km)	*Mean molecular mass corresponding to nitrogen in molecular form and oxygen in atomic form assumed*		*Mean molecular mass corresponding to nitrogen and oxygen in atomic form assumed*	
	Temperature (°K)	*Total particle concentration (/cm^3)*	*Temperature (°K)*	*Total particle concentration (/cm^3)*
180	560	$1{\cdot}2 \times 10^{10}$	340	$2{\cdot}0 \times 10^{10}$
200	655	$4{\cdot}3 \times 10^{9}$	395	$7{\cdot}1 \times 10^{9}$
220	750	$1{\cdot}8 \times 10^{9}$	455	$2{\cdot}9 \times 10^{9}$

VOL. 4 1953

a recent rocket flight (The Rocket Panel, 1952) suggest (cf Table 3). These measurements are preliminary and require confirmation. Use will not be made of them hereafter in order to minimize the risk of raising unreal objections to the acceptance of the altitudes deduced from the observations.

In spite of the efforts of a number of investigators no chemical reaction, other than three-body association, has yet been found which could excite the green line. The energy requirement for the red line is not so difficult to meet and both

$$N + O_3 \rightarrow NO_2 + O(^1D) \tag{8}$$

and

$$N + NO \rightarrow N_2 + O(^1D) \tag{9}$$

are exothermic. However (8) must be unimportant at great altitudes because of the rapid diminution of $n(O_3)$ above the $O - O_2$ transition layer; and a recent photo-chemical study (BATES, 1953) indicates that the distribution of nitric oxide is not such that (9) could lead to appreciable emission from the region favoured by BARBIER.

MITRA (1952) advocates the view that the airglow is due to the release of the ionization energy stored in the *F*-layer through the recombination of oppositely charged particles; he considers, for instance, that the forbidden lines of atomic oxygen might arise from

$$O^+ + O^- \rightarrow O + O(^1S \text{ or } ^1D), \tag{10}$$

or from some similar process. The predicted altitude of the emission is in general agreement with the altitudes deduced from the more recent measurements (cf Table 1). However the number of ions present in the layer at midnight is apparently at least several times smaller than the number required to maintain the observed photon intensity of $\lambda 5577$ until dawn (cf BATES and MASSEY, 1946). Though the discrepancy is not very large the quantities concerned seem to be determined too closely to allow it to be dismissed. Other difficulties have been raised elsewhere (BATES, 1948).

In the absence of any proper theory of their origin it is of course highly speculative to invoke fast electrons. Nevertheless the hypothesis that they provide excitation at great altitudes through inelastic encounters with oxygen atoms

$$O + e \rightarrow O(^1S \text{ or } ^1D) + e \tag{11}$$

must be considered. SEATON (1953), who has recently carried out a thorough quantal study of the processes, finds that the cross section associated with the excitation of the ^1S level is much less than that associated with the excitation of the ^1D level. From his detailed results it may readily be shown that the hypothetical fast electrons cannot be responsible for the green line; for, if they were abundant enough for this, it would appear that they would cause the red line to be at least about five times as intense as is observed.* The argument depends on the

* The possibility that the ^1D level alone is excited of course remains. SPITZER (1945) has suggested indeed that the thermal electrons in the high energy tail of the Maxwell distribution may be sufficient for the excitation, but it is not easy to reconcile this with the fact that the intensity of the red line remains almost constant throughout the sunspot cycle (cf RAYLEIGH and JONES (1935), DUFAY and TCHENG (1947a).

assumption that deactivation does not exert a major influence but, in spite of the fact that the radiative life of $O(^1D)$ is 110 sec (Garstang, 1951), there is little doubt that this assumption is valid. Thus it may be shown from Seaton's cross section data that if the electron concentration and temperature were $5 \times 10^5/cm^3$ and 2000°K respectively (both of which are high) the life of the metastable atoms to destruction by

$$O(^1D) + e \rightarrow O(^3P) + e \qquad (12)$$

would be about 1000 sec. Again, the effect of excitation transfer or chemical change in collisions with atoms or molecules can scarcely be important in view of the extremely low gas density. The only excitation transfer process which does not involve a large energy defect is indeed

$$O(^1D) + O_2(X\ ^3\Sigma_g^-, v = 0) \rightarrow O(^3P_2) + O_2(b\ ^1\Sigma_g^+, v = 2) - (0{\cdot}004 \text{ e volts}), \quad (13)$$

and this does not in fact lead to deactivation, for the reverse process, which also occurs, ($O_2(b\ ^1\Sigma_g^+)$ being metastable) is strongly favoured by the high $n(O)/n(O_2)$ ratio.* Destruction of the metastable atoms by some chemical reaction cannot be completely excluded but none of the obvious possibilities are at all plausible: for example even if the associated rate coefficient were as much as 10^{-10} cm^3/sec the reaction

$$O(^1D) + NO(X\ ^2\Pi) \rightarrow O_2(X\ ^3\Sigma_g^-) + N(^4S) + (0{\cdot}56 \text{ e volts}) \qquad (14)$$

would not be sufficiently effective unless $n(NO)$ near the 250 km level were at least some $10^8/cm^3$ which is most unlikely (Bates, 1953). Support for the predictions as to the influence fast electrons would have on the emission is furnished by the spectra of low latitude aurorae; for in these λ6300 is very much stronger than λ5577 (Götz, 1941, Dufay and Tcheng, 1942;) whereas in the spectrum of the normal airglow the two lines are of about equal intensity. It is suggestive, too, that λ5199 of atomic nitrogen is present in the former (Dufay, Gauzit and Tcheng, 1941) but is absent from the latter.

Another speculative possibility which may be mentioned is that the oxygen is excited by in-falling interplanetary atoms of velocity of perhaps 5×10^6 cm/sec (cf Ginzburg, 1952). The hypothesis cannot be rejected with the same assurance as can the hypothesis that sodium swept up by the Earth is *itself* excited in collisions with the atmospheric constituents. However it would scarcely be anticipated that the intensity would be sufficient; and again the pronounced *seasonal* variation (Rayleigh and Jones, 1935) would be difficult to understand with such a source.

Finally it may be pointed out that *irrespective of the excitation processes* the lack of correlation between the variations in the intensities of λ5577 and λ6300 (cf Dufay and Tcheng, 1947a) indicates that there must be significant deactivation;† and as has been seen it is not easy to understand how this could occur at great altitudes.

* It is to be noted also that difficulties would arise in connection with polar aurorae if the mechanism supposed to be operative in suppressing the red line at great altitudes were of such a nature that it would be very much more effective at the 100 km level.

† The lower level of the green line is of course the upper level of the red line.

VOL. 4 1953

The Herzberg and Vegard-Kaplan band systems

Arguments similar to those advanced in the previous section make it exceedingly unlikely that the Herzberg band system is excited by processes involving charged particles, such as

$$O_2^+ + O^- \rightarrow O_2(A\ ^3\Sigma_u^+) + O, \tag{15}$$

or

$$O_2 + e \rightarrow O_2(A\ ^3\Sigma_u^+) + e; \tag{16}$$

and a very powerful additional argument is provided by the fact that the system is not enhanced in aurorae. As far as is known simple three-body association

$$O + O + X \rightarrow O_2(A\ ^3\Sigma_u^+) + X \tag{17}$$

is the only suitable chemical process. This should be rapid enough to yield the required photon intensity (some 2×10^8/(cm^2 column sec) BARBIER, 1953) from the region near the 100 km level, but for the reasons detailed in the discussion on the related process (7) it could not give rise to appreciable emission from great altitudes. Further elaboration of the theoretical considerations is unnecessary for the contention that the luminous layer lies at an altitude of 200–250 km is in conflict with certain measurements. From studies of the profiles of the bands SWINGS (1943) deduced that the rotational temperature of the emitting oxygen molecules is about 150°K, and BARBIER (1947c) that it is between 170 and 220°K.* Such temperatures almost certainly do not occur above about the 100 km level—and if in fact they did occur the gas density at altitudes of 200–250 km would be much lower than hitherto assumed, and the already cogent case against there being any appreciable emission from this region would be correspondingly strengthened.

Much the same considerations apply to the Vegard-Kaplan band system and we content ourselves with recalling that according to CABANNES and DUFAY (1946) the rotational temperature is only 230°K, which is much lower than would be expected if the emission originated from near the 350 km level as claimed.

Meinel band system

The altitude of 70 km favoured by BARBIER (1952) for the Meinel band system is consistent with the rotational temperature being 260°K, as the direct measurements indicate (MEINEL, 1950), and is in harmony with the views of the theorists (BATES and NICOLET, 1950b), so that a detailed discussion would be superfluous. This is the only case in which there is not a most grievous discrepancy and it is at least suggestive that the system concerned is the one most suited to observational work,† the corrections for atmospheric extinction and for any extra terrestrial component being comparatively small owing to its spectral position and its great intrinsic intensity.

* These results admittedly cannot be regarded as final since the band structure was not resolved (NICOLET, 1952).

† As evidence of the difficulty of the observational work even here it may be mentioned that HURUHATA (1950) obtained an altitude of 300 km. BARBIER rejects this result as he considers that ROACH, PETTIT and WILLIAMS (1950), who obtained an altitude of 70 km, took greater precaution in eliminating systematic errors.

*λ*5180 *continuum*

The processes which might be invoked to explain the non-discrete radiation are limited to free-free, bound-free and free-bound transitions. As will be seen none of these appears capable of emitting a continuum from the 460 km region whose photon intensity is as much as the observed 8×10^5/(Å cm^2 column sec) (Barbier, Dufay and Williams, 1951).

Free-free transitions may be ignored from the onset: for the photon energy in the continuum under discussion is as high as 2·4 e volts and the number of particles with this much kinetic energy is utterly inadequate.

Several arguments may be adduced for dismissing bound-free transitions, but the following suffices. Nitrogen is the only molecular species which is not undoubtedly exceedingly rare at an altitude of 460 km; as its dissociation energy is 7·4 e volts or more (Herzberg, 1950; Gaydon, 1953) the excitation energy of the unknown bound level concerned would have to be at least 10 e volts; so much energy could only be provided through collisions with fast particles; and sufficient particles with 10 e volts kinetic energy are not normally present since if they were they would indubitably excite many of the lines and bands characteristic of aurorae.

Free-bound transitions involving electrons need not be discussed at length. There are two principal types, radiative recombination and radiative attachment. It has long been established that the former is unimportant under upper atmospheric conditions (Bates, Buckingham, Massey and Unwin, 1939). As regards the latter, though

$$\mathrm{O} + e \rightarrow \mathrm{O}^- + h\nu \tag{18}$$

may perhaps give a weak but observable continuum* the centre of the luminosity, far from being in the 460 km region, must be somewhere between the E and F-layers; for the photon yield from any level is proportional to $\eta n(\mathrm{O})n(e)$, and η, the attachment coefficient, is known from quantal investigations (Bates and Massey, 1943) to be either approximately constant or a *decreasing* function of the temperature.

For atomic free-bound transitions such as

$$\mathrm{O} + \mathrm{O} \rightarrow \mathrm{O}_2 + h\nu \tag{19}$$

it is conceivable (though unlikely) that the rate coefficient increases so rapidly with the temperature that the main emission takes place in the 460 km region in spite of the fall off with altitude of the gas density. However, if this should be the case the usual difficulties would arise in connection with the absolute intensity. At temperature T the differential rate coefficient for the photon yield from atoms whose nuclear separation lies between R and $R + \mathrm{d}R$ is given by

$$\gamma(T, R)\mathrm{d}R = \frac{256\pi^5}{3h\lambda(R)^3} D(R)^2 r \exp\{-U(R)/kT\} R^2 \mathrm{d}R \tag{20}$$

* This continuum probably lies in the green or yellow but its spectral position cannot be predicted closely as the electron affinity of atomic oxygen has not yet been determined with precision (cf Massey, 1950).

VOL. 4 195

where h is Planck's and k Boltzmann's constant, λ is the wavelength of the radiation emitted, $D(R)$ is the dipole moment associated with the electronic transition involved, $U(R)$ is the potential energy of the initial state (measured with respect to the dissociation limit), and r is the ratio of the statistical weight of this state to the total statistical weight of all possible states of the colliding system (BATES, 1951b). On taking λ to be 5180 Å and substituting numerically (20) becomes

$$\gamma(T, R)\,\mathrm{d}R = 1{\cdot}7 \times 10^{8} D(R)^2 r \exp\{-U(R)/kT\} R^2\,\mathrm{d}R \text{ cm}^3/\text{sec} \qquad (21)$$

where R is in cm and $D(R)$ is in atomic units. To favour the hypothesis under consideration it will now be assumed arbitrarily that the $^3\Pi_u$ potential curve is of a suitable form and that (19) is brought about through the $^3\Pi_u - {}^3\Sigma_g^-$ transition. Like all others which might be invoked this is of the sub-Rydberg type (cf HERZBERG, 1950) so that $D(R)$ can scarcely exceed unity and is probably much less. Inspection of the $^3\Sigma_g^-$ potential curve indicates that $R(5180A)^2$ is about 3×10^{-16} cm^2 and that a $\mathrm{d}R$ of about 1×10^{-12} cm corresponds to a 1 Å wavelength spread. Since r is 6/81 it follows from (21) that R the rate of emission of photons is at most some

$$\{4 \times 10^{-21} \exp(-U(R)/kT) n(\mathrm{O})^2\}/(\text{Å cm}^3 \text{ sec}) \qquad (22)$$

The exponential factor at the 460 km level would have to be very small for P there to be even 4 times P at, say, the 200 km level. For example it would have to be $2{\cdot}5 \times 10^{-3}$ if, to take a rather extreme distribution

$$n(\mathrm{O}|460 \text{ km})/n(\mathrm{O}|200 \text{ km}) = 0{\cdot}1, \qquad T(460 \text{ km})/T(200 \text{ km}) = 2. \qquad (23)$$

With this value the limit in (22) reduces to about

$$1 \times 10^{-23} n(\mathrm{O})^2/(\text{Å cm}^3 \text{ sec}) \qquad (24)$$

which clearly cannot be reconciled with the observed intensity. Because of the possibility that $n(\mathrm{N}|460 \text{ km})/n(\mathrm{N}|200 \text{ km})$ may be larger than the corresponding ratio for atomic oxygen the case against accepting

$$\mathrm{N} + \mathrm{O} \rightarrow \mathrm{NO} + h\nu, \qquad (25)$$

though strong, is not absolutely conclusive.

Quite apart from the difficulties in attributing the luminosity to the 460 km region it must be remembered that some of the processes occurring in the 100 km region would be expected to yield an intense continuum. For instance the fact that the temperature is but moderate does not prevent (19) from proceeding through the $A\,^3\Sigma_u^+ - X\,^3\Sigma_g^-$ transition since the $A\,^3\Sigma_u^+$ potential curve is attractive; and from figures given elsewhere (BATES, 1951b) it may be seen that to provide the photon intensity which is observed, 8×10^5/(Å cm^2 column sec), the associated differential rate coefficient need only be of order 10^{-26} cm^3/(Å sec).

REFERENCES

ABADIE, P., VASSY, A. and VASSY, E. 1949 Ann. Géophys. **5**, 157

BARBIER, D. 1944 Ann. Géophys. **1**, 144

Barbier, D.	1947a	Relations entre les Phénoménes Solaires et Géophysiques, p. 182 (Paris: Centre National de la Recherche Scientifique)
	1947b	Ann. Astrophys. **10**, 141
	1947c	Comptes Rendus (Paris) **224**, 635
	1952	Mém Soc. Roy. Sciences Liège **12**, 43
	1953	Ann. Astrophys. **16**, 96
Barbier, D., Dufay, J. and Williams, D.	1951	Ann. Astrophys. **14**, 399
Barbier, D., and Pettit, H.	1952	Ann Géophys. **8**, 232
Bates, D. R.	1947	Terr. Mag. Atmos. Elect. **52**, 71
	1948	Emission Spectra of the Night Sky and Aurora, p. 21 (London: Physical Society)
	1951a	Proc. Phys. Soc. B. **64**, 805
	1951b	Mon. Not. Roy. Astr. Soc. **111**, 303
	1953	The Solar System: Vol. 2. The Earth as a Planet. (Ed. G. P. Kuiper) (Chicago: University Press)
Bates, D. R., Buckingham, R. A., Massey, H. S. W. and Unwin, J. J.	1939	Proc. Roy. Soc. A. **170**, 322
Bates, D. R. and Massey, H. S. W.	1943	Phil. Trans. Roy. Soc. A. **239**, 269
	1946	Proc. Roy. Soc. A. **187**, 261
Bates, D. R. and Nicolet, M.	1950a	J. Geophys. Res. **55**, 235
	1950b	J. Geophys. Res. **55**, 301
Bricard, J. and Kastler, A.	1944	Ann. Géophys. **1**, 53
Cabannes, J. and Dufay, J.	1946	Ann. Géophys. **2**, 290
Chamberlain, J.W. and Oliver, N. J.	1953	Phys. Rev. **90**, 1118
Dufay, J., Gauzit, J. and Tcheng, Mao-Lin	1941	Cahiers de Physique, **1**, 59
Dufay, J. and Tcheng, Mao-Lin	1942	Cahiers de Physique **8**, 2
	1947a	Ann. Géophys. **3**, 153
	1947b	Ann. Géophys. **3**, 282
Garstang, R. H.	1951	Mon. Not. Roy. Astr. Soc. **111**, 115
Gaydon, A. G.	1953	Dissociation Energies, 2nd Ed. (London: Chapman and Hall)
Ginzburg, V. L.	1952	Dok. Akad. Nauk. **84**, 245
Götz, F. W. P.	1941	Naturwiss. **29**, 690
Herlofson, N.	1947	Reports on Progress in Physics **11**, 444
Herzberg, G.	1950	Spectra of Diatomic Molecules, 2nd Ed. (London: MacMillan and Co.)
Huruhata, M.	1950	Tokyo Astr. Obs. Repr. No. 71
Kahn, F. D.	1950	Mon. Not. Roy. Astr. Soc. **110**, 483
Massey, H. S. W.	1950	Negative Ions, 2nd Ed. (Cambridge: University Press)
Massey, H. S. W. and Burhop, E. H. S.	1952	Electronic and Ionic Impact Phenomena (Oxford: Clarendon Press)
Meinel, A. B.	1950	Astrophys. J. **112**, 120
Mitra, S. K.	1952	The Upper Atmosphere, 2nd Ed. (Calcutta: Royal Asiatic Society of Bengal)
Nicolet, M.	1952	Mém. Soc. Roy. Sciences Liège **12**, 71
Rayleigh, Lord and Jones, H. Spencer	1935	Proc. Roy. Soc. A. **151**, 22
Roach, F. E. and Barbier, D.	1950	Trans. Amer. Geophys. Union **31**, 7

VOL. 4 1953

ROACH, F. E. and PETTIT, H.	1951	J. Geophys. Res. **56**, 325
	1952	Mém. Soc. Roy. Sciences Liège **12**, 13
ROACH, F. E., PETTIT, H. and WILLIAMS, D. R.	1950	J. Geophys. Res. **55**, 183
SAHA, A. K.	1952	Mém. Soc. Roy. Sciences Liège **12**, 97
SANFORD, R. F.	1950	Publ. Astr. Soc. Pacific **62**, 272
SEATON, M. J.	1953	Phil. Trans. Roy. Soc. A. **245**, 269
SPITZER, L.	1948	The Atmospheres of the Earth and Planets (Ed. G. P. Kuiper) p. 213. Chicago: University Press).
SWINGS, P.	1943	Astrophys. J. **97**, 72
The Rocket Panel	1952	Phys. Rev., **88**, 1027
VAN RHIJN, P. J.	1921	Publ. Astr. Lab. Groninger, No. 21

OL.
4
953

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 124 to 128. Pergamon Press Ltd., London

The agrimeter for continuous recording of the atmospheric electric field

J. Alan Chalmers

Senior Lecturer in Physics, Durham Colleges in the University of Durham

(*Received* 17 *July* 1953)

ABSTRACT

A new apparatus is described for giving continuous records of the atmospheric electric field at the earth's surface, suitable for use in various conditions. Some typical records are given.

1. Introduction

Various methods have been used for the measurement of the vertical electric field in the atmosphere close to the earth's surface; these can be divided into two main groups, based respectively on the potential at a point in the atmosphere and on the "bound" charge on a portion of the earth's surface.

The present work describes an apparatus of the second group, designed so as to be able to function over a wide range of sensitivities and under all conditions of weather.

2. Principle of the Agrimeter

The present apparatus has been called the "Agrimeter" (derivation obvious!) to distinguish it from the more familiar "field mill."

The action is based on the following principle: a fixed earthed cover has an opening under which a moving plate can pass. The plate is connected to earth when it is beneath the opening and its upper surface then carries the charge corresponding to the lines of force ending on the plate, a charge proportional to the field strength. The earth connection is broken while the plate is exposed to the maximum number of lines of force, and the plate carries its charge with it. When the plate has passed under the cover, it can make contact with a "collector" and gives up a proportion of its charge. If the field strength remains constant, each time the plate reaches the collector it carries the same amount of charge and it is at the same potential. Measurement is then made either of the charge or of the potential.

3. Construction of the Agrimeter

An electric motor of $\frac{1}{4}$ H.P., giving 1425 revs/min drives a horizontal steel shaft, mounted on ball bearings, through a rubber coupling to avoid transmission of vibration. On the shaft are mounted, at a distance apart of 30 cm, two *Tufnol* discs of thickness 1 cm and radius 10 cm. Six copper plates of thickness 4 mm and breadth 6 cm are fixed, parallel to the shaft, at equal intervals on the rims of the discs, forming a kind of "squirrel cage"; Fig. 1 shows the apparatus with the motor at rest. Fig. 2 shows the end-on view at the end remote from the motor; F is one disc and B · · · B the copper plates.

G is another, smaller *Tufnol* disc of radius 4 cm and thickness 1·5 cm. G is in contact with F and so the surface of G in Fig. 2 is 1·5 cm in front of that of F.

VOL.
4
1953

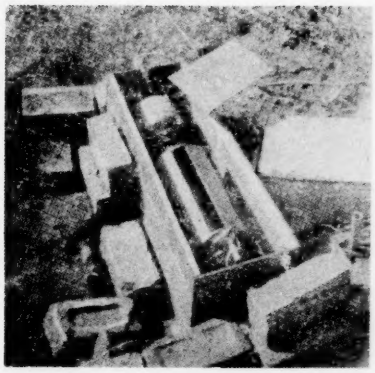


Fig. 1. Agrimeter open and at rest

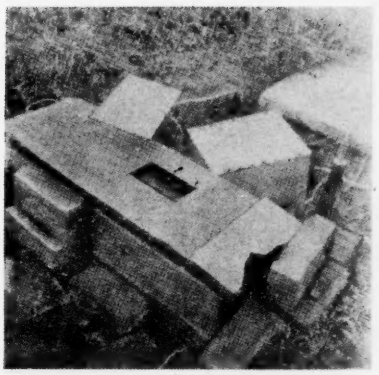


Fig. 3. Agrimeter closed and in use

OL.
4
953

Fig. 4. 17-hour record in fine weather

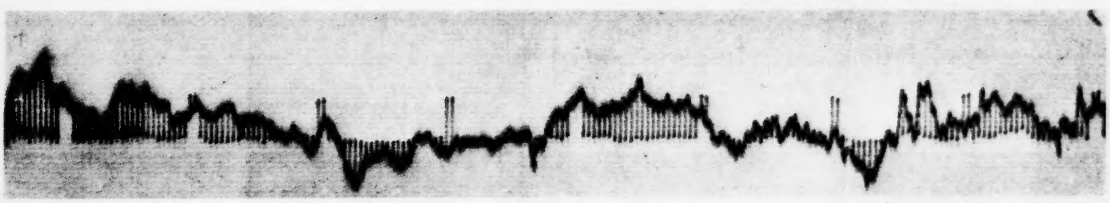


Fig. 5. 17-hour record in disturbed weather

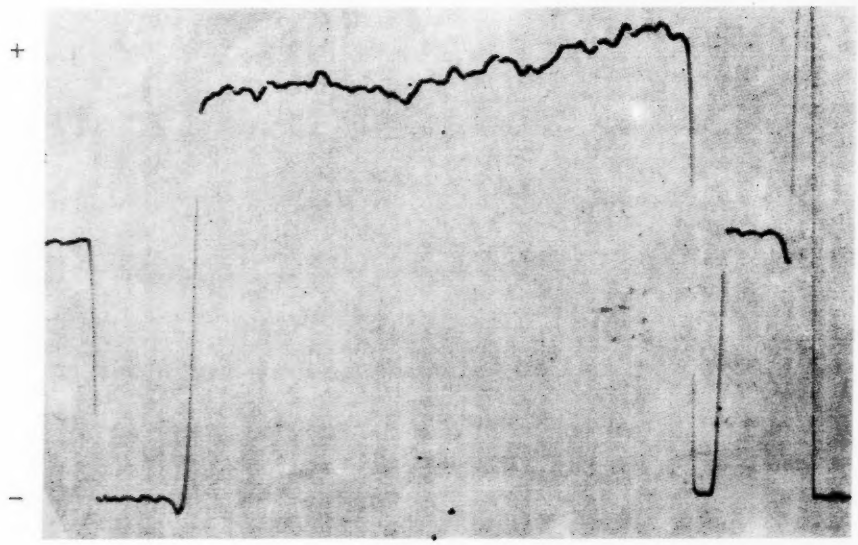


Fig. 6. 13-minute record in fine weather with open agrimeter

VOL.
4
1953

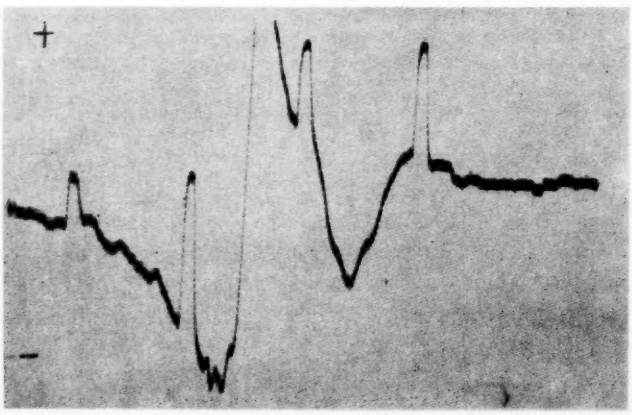


Fig. 7. 28-minute record in shower

On the rim of G are fixed 12 brass sectors A · · · A, with intervals of about 1 mm between them; the radial thickness of the sectors A · · · A is 2 cm. The sectors A · · · A are turned so that the outer surfaces all lie exactly on a cylindrical surface concentric with the shaft. Alternate brass sectors are connected to corresponding copper plates, the intermediate sectors being left unconnected.

Two spring-loaded carbon brushes, C and D, are fitted so as to make contacts with the sectors A, C when the corresponding copper plate B is at its highest position and D when it has moved round through about 60°.

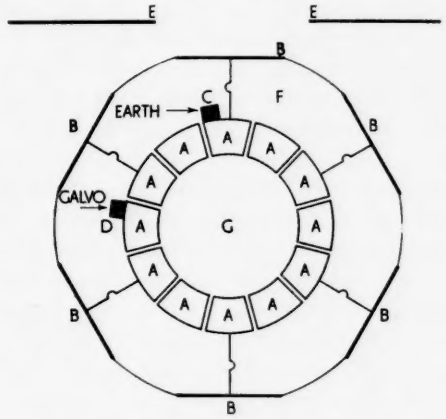


Fig. 2. Diagrammatic end-on view.

The whole is mounted on a steel base and surrounded on four sides by a wooden box, on the top of which is placed a copper lid E, with a hole 17 cm by 9 cm. The hole is arranged to be placed symmetrically above the middle of the cage. The lid is then able to prevent rain from reaching the *Tufnol* discs, and any splashing from the plates will take the water to the box rather than to the discs. The whole apparatus is sunk into the ground so that the lid E is approximately in the plane of the earth's surface. Fig. 3 shows the agrimeter closed, in use.

The upper brush, C, is connected to earth and to an earth line into the building housing the measuring instruments. The lower brush, D, is connected by coaxial cable to the measuring instrument, the other terminal of which is connected to the earth line. The motor, the copper lid, the sheath of the cable and the steel base are connected to earth and an earthed phosphor bronze spring strip makes contact with the shaft; there is also a connection across the rubber drive.

4. Method of measurement

The output from the collector can be used in two distinct ways. If the collector is connected directly to an insulated instrument for measuring potential, such as a quadrant electrometer, then, in a steady field, the electrometer acquires by stages the potential of the plate at the time of contact with the collector; if the field alters, the altered potential of the moving plate is communicated to the electrometer after a number of contacts which depends on the relative capacities of the

plate and the electrometer system. This was the method used by RUSSELTVEDT (1925), WORKMAN and HOLZER (1939) and GOTO (1951), in earlier forms of the apparatus; in each case an electrometer of small capacity was used, so that a rapid response was obtained. This method has the advantage of being independent of the motor speed.

On the other hand, if the collector is connected directly to earth through a galvanometer of low resistance, then the charge on the plate each time it makes contact with the collector runs directly through the galvanometer. There is then an intermittent current through the galvanometer, and if the interval between successive contacts is small compared with the time of oscillation of the galvanometer, there will be a steady deflection in a constant field and the response to field changes will be governed by the response of the galvanometer. This method has the great advantage that it does not depend upon very high insulation in the agrimeter; so long as the insulation resistance remains high compared with the galvanometer resistance, the current gives the correct measure of the field. Another advantage, compared with the potential method of measurement, is the ease with which the sensitivity can be altered by alteration of the galvanometer shunt. In order to minimize effects due to thermoelectric and other potentials, it is found useful to insert a resistance of some megohms between the agrimeter and the galvanometer. Variation of motor speed, due to variation of mains voltage, is eliminated by the use of a constant voltage transformer.

In fine weather, it is possible to increase the output by removing the copper lid, thus exposing a much larger area of the plates to the field.

VOL. 4 1953

5. THEORY OF THE AGRIMETER

A plate of area A cm^2 exposed to a field of strength F carries a charge

$$Q = FA/4\pi$$

when Q and F are expressed in E.S.U. If Q is in coulombs and F in volts/metre,

$$Q = FA \times 10^{-13}/36\pi$$

If the capacity of the plate when in contact with the collector is C, the capacity of the collector, cable and connected apparatus D, the resistance between the cable and earth R, and the time between successive contacts of plates and collector T, then it is simple to show that the average current through the resistance R is:

$$I_R = \frac{Q}{CR + T}$$

provided that C and T/R are small compared with D.

The potential developed across R is:

$$V_R = \frac{QR}{CR + T}$$

The two types of measurement discussed above consist of measuring V, with R very large, giving:

$$V_\infty = Q/C$$

and measuring I, with R small, giving

$$I_0 = Q/T.$$

Measurements, in addition, have been made with $R = 94$ megohms and the results have been consistent with the theory, giving a value of C of about 31 $\mu\mu$f.

For calibration, a plate is arranged to move over the opening at definite intervals, at times carrying zero potential and at times a fixed potential. Absolute calibration was carried out by comparison with the potential on a stretched wire carrying a glowing fuse.

The galvanometer used has a sensitivity of 1 cm $= 6{\cdot}32 \times 10^{-4}$ μa and a field of 100 v/m gave a deflection, with the agrimeter closed, of 0·53 cm. With the agrimeter open, the deflection was 4·8 cm.

Calculations show that rain currents are not able to affect the galvanometer readings appreciably.

6. Typical records

Figs. 4 and 5 show 17-hour records in fine and disturbed weather respectively; the earthed movable plate covers the hole for ½ minute every 5 minutes, but every 2 hours the plate is charged for 10 minutes to 12·2 volts, giving a field corresponding to 277 v/m.

Fig. 6 shows a 13-minute record in fine weather with the agrimeter open. At the start and finish of the record are periods with a plate over, with fields corresponding to 0 and 133 v/m. The vertical white lines represent intervals of ½ minute. Fig. 7 shows a 28-minute record during a shower, with the agrimeter closed and at a low sensitivity. At 3 of the 5-minute intervals, the covering plate is earthed and at the other two it carries a potential of 60 volts corresponding to 1300 v/m.

7. Performance

The agrimeter has been run continuously for several months without attention beyond occasional tightening of the spring loading of the contacts; when the springs are too loose, the output shows an oscillation of period around ½ minute, as can be detected on some of the records. After some months' use it has been found necessary to adjust bearings and replace brushes, but otherwise the apparatus has needed little care.

Even in snow, or very heavy rair, there has been no indication that the agrimeter did not behave normally.

Small drifts and irregularities of the zero have been noticed, but, with the regular calibration, these are not serious.

Attempts have been made to feed the potential output into an amplifier to provide compensation for field changes, in the manner used by Goto (1951), but the output has not been steady enough for these attempts to have yet been successful; it may be that an improved brush system will be required.

8. Acknowledgments

The author wishes to thank:

Mr. J. Johnson, technician of the Physics Department, who constructed the agrimeter and the covering device.

Dr. W. W. Mapleson and Mr. J. Stockill, research students, who constructed the timing devices and assisted with calibrations.

The Council of the Durham Colleges for the grant of a sabbatical year which has made it possible for the present work to be carried out without interruption from other duties; and for grants from their Research fund for apparatus and materials used.

References

Goto, M.	1951	J. Geomag. and Geoelec. **3**, 22
Russeltvedt, N.	1925	Beiheft zum Jahrbuch des Norwegischen Meteorologischen Instituts für 1925
Workman, E. J. and Holzer, R. E.	1939	Rev. Sci. Inst. **10**, 160

VOL.
4
195

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 129 to 140. Pergamon Press Ltd., London

The reception of atmospherics at high frequencies

F. Horner, M.Sc., A.M.I.E.E.

D.S.I.R., Radio Research Station, Slough

(*Received* 18 *February* 1953)

ABSTRACT

Observation of the amplitudes of atmospherics received in the high frequency band offers a possible method of assessing the distances of their sources. In a preliminary investigation of the potentialities of the technique, atmospherics have been recorded simultaneously at two frequencies in the upper part of the h.f. band. Reception would be expected at both frequencies from sources within ground-wave range (about 100 km). At rather greater distances, such that the receiving point is in the skip zone for both frequencies, no reception should be possible. There should then be a zone from which there should be reception on the lower frequency only, and at greater distances energy should again be received at both frequencies.

The preliminary results conform generally with this pattern, and the boundaries of the various zones are in agreement with the ionospheric data derived from vertical incidence soundings. The sources of the atmospherics observed were located by use of the network of direction finders operated by the Meteorological Office. It is concluded that the technique forms a basis for routine estimation of the distances of storms and merits further examination to determine more precisely the accuracy which can be achieved and to overcome certain practical difficulties in the observations.

1. Introduction

Investigations into the nature of atmospherics and their propagation over the earth have hitherto been conducted mainly at frequencies below 100 kc/s—a part of the spectrum in which the energy radiated by an atmospheric is high. At these frequencies it is not difficult to study individual atmospherics for they have large amplitudes and are separated by periods of low noise field intensity. Examples of past developments are the location of thunderstorms by direction finding on atmospherics, improvement of knowledge of the characteristics of the discharge, and the recording of waveforms of atmospheric to study propagation phenomena at very low frequencies—in particular, the properties of the ionosphere.

As the frequency is raised, the study of atmospherics becomes increasingly difficult, partly because their intensity diminishes and becomes comparable with the large number of radio signals which almost fill the spectrum, and partly because many more atmospherics are received and they tend to overlap and produce continuous noise. However there is a need for a method of assessing the distances of sources of atmospherics by observations at a single point and the possibilities of using measurements at high frequencies should not be overlooked. One possible method of location involves the determination of direction by the usual methods at low frequencies, coupled with the determination of the maximum frequency in the h.f. band at which energy is received from the atmospheric. The skip distance corresponding to this frequency should be calculable from a knowledge of ionospheric conditions.

In this paper a preliminary investigation of the possibilities of this technique is described. The main objectives were to determine whether individual atmospherics could be distinguished in the general noise at high frequencies, to relate the reception of atmospherics to the ionospheric data and to determine the accuracy required for an operational system to be practical.

2. Observation Technique

Atmospherics were recorded simultaneously at two frequencies which were normally 1 to 3 Mc/s apart in the upper portion of the h.f. band. The principle of the apparatus is illustrated in Fig. 1. Two receivers of equal gain and bandwidth (10 kc/s) were connected to a common aerial and the rectified output voltages

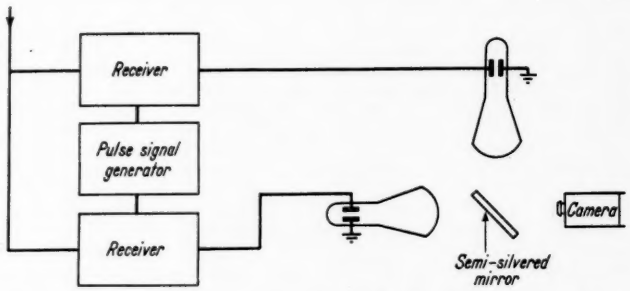


Fig. 1. Apparatus for recording atmospherics at high frequencies.

applied to two cathode ray tubes. These were mounted in such a way that by using a semi-reflecting mirror both traces could be photographed on a single 35 mm film moving at uniform speed (0·5 cm/sec). No time bases were applied to the tubes, the receiver outputs simply deflecting the spots in directions normal to the direction of travel of the film. A sample record is shown in Fig. 2, in which the output voltages of the receivers are represented by deflections from the centre of the film; the reference marks along the lower edge of the film were made by lighting a neon lamp in the camera by pressing a switch. The marks, (a), (b), (c), (d), refer to atmospherics observed by cathode-ray direction-finders (C.R.D.F.) from sources marked correspondingly in Fig. 8. Only two of these atmospherics were received at high frequencies.

The gains of the receivers could be equalised either by applying a signal from a generator or by tuning to a signal at some frequency between those to be used. These methods gave mutually consistent results.

The experiments were carried out in collaboration with the network of atmospherics cathode ray direction finders operated by the Meteorological Office (Clarke and Mortimer, 1951). When storms were located in suitable situations the frequencies of the two receivers were adjusted, by reference to predictions of ionospheric conditions, on either side of the expected cut-off frequency for the distance involved.

The camera was then set in motion and the neon lamp triggered each time an atmospheric was selected by the C.R.D.F. operators. Fig. 2 illustrates the time delay, about $\frac{1}{2}$ second, involved in "calling" the atmospheric and triggering the neon lamp. Each sequence of observations extended over a few minutes during the normal period of routine observations of the C.R.D.F. network. In this paper such a sequence will be called a "run."

VOL. 4 1953

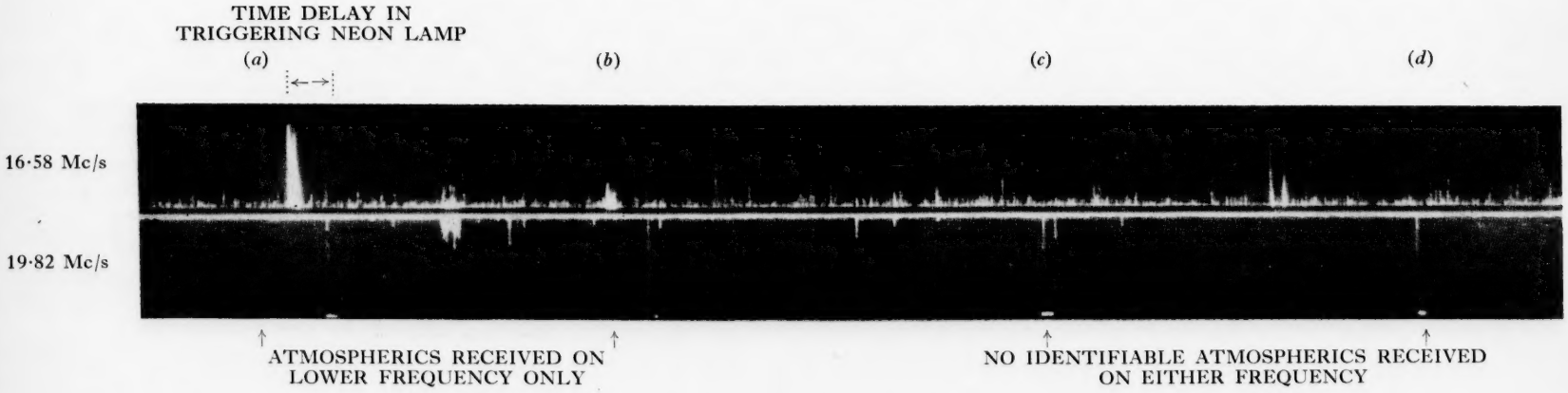


Fig. 2. Record obtained in good conditions (1500 hrs on 14 December, 1950). Sources are plotted in Fig. 8

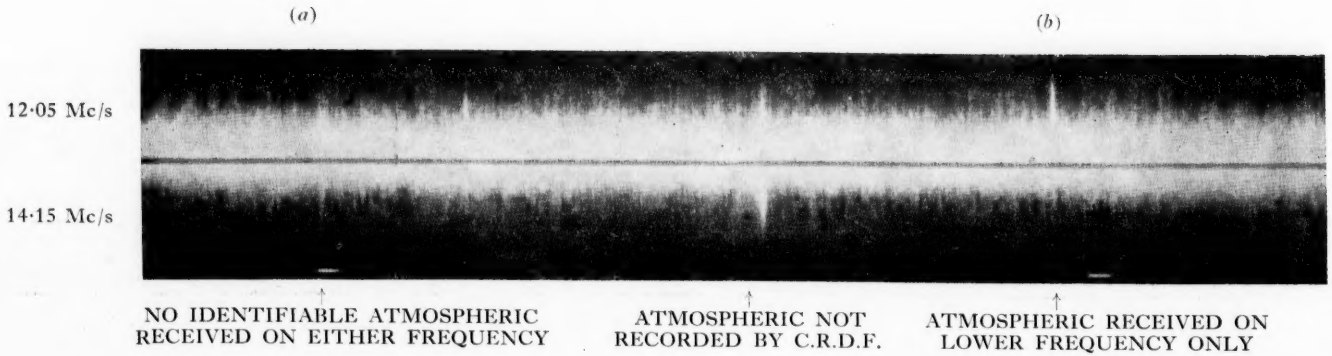


Fig. 3. Record obtained during severe interference (1500 hrs on 18 December, 1950). Sources are plotted in Fig. 7

VOL.
4
1953

All the observations were made at Teddington, Middlesex, during daylight hours, at which times the ionospheric conditions were relatively stable. Considerable man-made interference was experienced and identification of corresponding atmospherics at low and high frequencies would have been much easier under good conditions. However it was often possible to recognise atmospherics through the interference even when this was severe. An example is presented in Fig. 3; this shows part of the record corresponding to Fig. 7 in which two sources (a), (b), are marked. A trace from (b) was received clearly at the lower frequency while nothing was received from (a). When the interference consisted of more isolated pulses, identification of the atmospherics was often difficult, but they were usually of longer duration than other noise impulses. Identification was most difficult when two or more atmospherics were received in rapid succession; it was necessary to ignore some observations which were ambiguous in this respect.

The success of the technique depends on the energy radiated by the source being uniform over the band in which the observations were made. The results so far indicate that the spectrum is sufficiently uniform for the purpose.

3. Comparison with Ionospheric Data

Let us suppose that only one reflecting layer affects the propagation of the atmospherics at the frequencies of observation. The area around the receiving point may then be divided into four zones:

(*a*) A ground-wave zone, from which energy is received by direct radiation from the source (radius about 100 km).

(*b*) A "skip zone," from which no energy should be received at either frequency.

(*c*) An "intermediate zone" from which energy should be received only at the lower frequency.

(*d*) An "outer zone" from which there should be reception at both frequencies, provided that the ionospheric absorption is not too great.

In most of the experiments described, the *F* layer controlled the reception, but in some instances it was necessary to consider sporadic *E*-ionisation.

The results are presented in the form of charts on which the sources of atmospherics are plotted in the positions indicated by the C.R.D.F. network. These positions are subject to some uncertainty, due to errors in direction finding, and the range errors at ranges greater than 1500 km may be serious. The plotted points are divided broadly into three categories according to whether energy was received on neither frequency, on the lower frequency only or on both frequencies. Again there is some uncertainty since the traces were being examined in the presence of man-made interference and other atmospherics. However in general an atmospheric is regarded as received on one frequency only when the amplitude appears to be at least ten times that on the other frequency. Frequently an atmospheric was received strongly on the lower frequency with a small trace on the upper frequency. This condition was probably due to scatter from the ionosphere when the upper frequency was slightly above the critical value. On a few occasions a detectable atmospheric was received only at the higher frequency, but only one such case occurs in the plotted charts (Fig. 9).

4. Results

4.1. *Sources within ground-wave range*

During local storms atmospherics were received by direct radiation from distances up to 100 km. They were characterised by large amplitudes and the build-up and decay of the atmospheric could usually be observed. Fig. 4 shows three examples of atmospherics from a distance of 70 km. In Fig. 4(*a*) the amplitudes are nearly equal, but in the other two examples, obtained one hour later, a greater signal

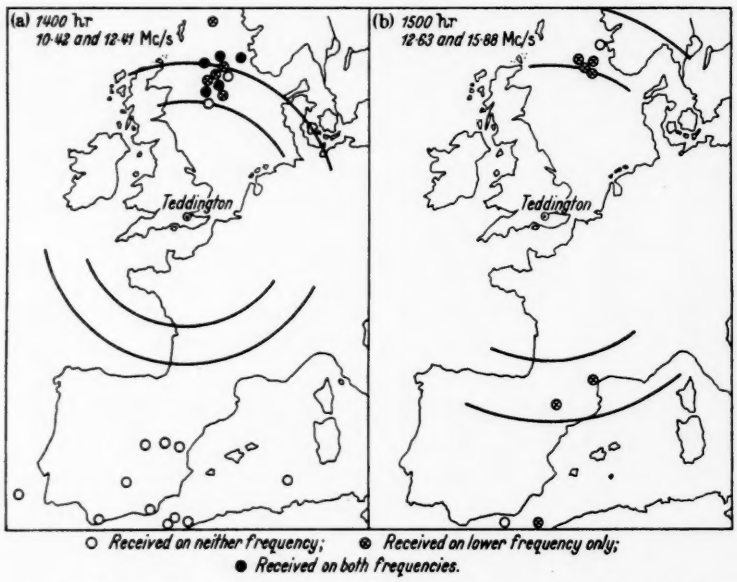

(a) 1400 hr
10·42 and 12·41 Mc/s
Teddington
(b) 1500 hr
12·63 and 15·88 Mc/s
Teddington
○ Received on neither frequency; ⊗ Received on lower frequency only;
● Received on both frequencies.

Fig. 5. Plots showing a storm centre in the intermediate zone, 10th October, 1950.
(*a*). 1400 hrs; 10·42 and 12·41 Mc/s.
(*b*). 1500 hrs; 12·63 and 15·88 Mc/s.

was received on the lower frequency. This may have been due to the propagation of energy by way of the ionosphere, since the frequencies were just below the vertical incidence critical frequency.

Only a few atmospherics received by ground-wave propagation have been recorded; most of these were similar to that illustrated in Fig. 4(*a*).

4.2. *Sources within the skip zone*

On one run (1500 hrs on 17 July 1950) frequencies of 12·32 and 15·06 Mc/s were used, corresponding to skip distances of 1150 km and 1520 km estimated from Slough (England) vertical incidence data. There were widespread storms in the north of England and 32 fixes obtained by C.R.D.F. were well within the skip zone. Of these, only one coincided with a trace recorded at high frequencies and there was some doubt whether this was a true correlation.

On a number of other occasions some sources have been plotted in the skip zone and in general have not been detected on either frequency.

VOL. 4 1953

(a)

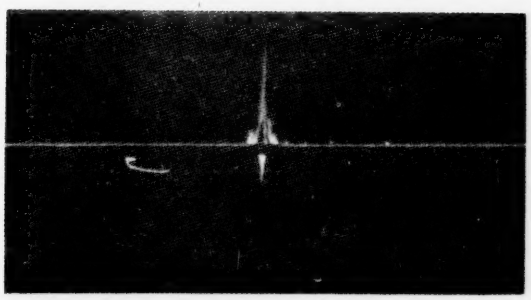
9·75 Mc/s

14·0 Mc/s

(b)

10·0 Mc/s

14·0 Mc/s



(c)

10·0 Mc/s

14·0 Mc/s

Fig. 4. Atmospherics received over a ground path of 70 km.

VOL.
4
1953

4.3. *Sources in or near the intermediate zone*

The most interesting results were obtained when storms were in the neighbourhood of the intermediate zone, and frequencies were usually chosen to achieve this condition.

Fig. 5 shows plots of two runs taken at an interval of one hour; the arcs representing the skip distances are plotted from Slough ionospheric data with

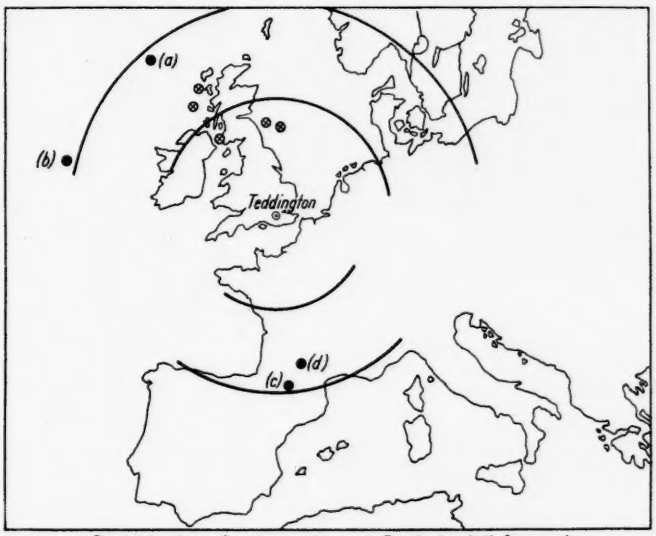


Fig. 6. Plot showing a storm centre at the skip distance. 1700 hrs on 1 May 1950. Frequencies 10·22 and 13·90 Mc/s.

small corrections for latitude variation. In Fig. 5(*a*) traces received from the North Sea were in all categories, which may be regarded as confirmation that the storm was in the region of the intermediate zone. It is not possible to say whether the mingling of the sources in different categories was due to errors in C.R.D.F. plotting or to random variations in the reflecting properties of the ionosphere—probably both factors were present. The more distant sources observed in this run will be discussed later.

For the run one hour later (Fig. 5(*b*)) the frequencies were changed. Only four atmospherics received from the North Sea storm corresponded with C.R.D.F. fixes and these again supported the placing of the storm in the intermediate zone. Agreement was also obtained on two sources to the south; the source at a greater distance to the south does not agree, but the C.R.D.F. fix was poor.

Another run is plotted in Fig. 6. This plot suggests that either the range was underestimated by C.R.D.F. or that the skip distances were less than those assumed. The sources marked (a), (b), (c) and (d) are discussed further in the next section.

Results for sources at greater ranges illustrate how observations at high frequencies may supplement those obtained by C.R.D.F. techniques. At long ranges

direction finding errors lead to serious uncertainties in range. The plot of a set of four bearings may be regarded as defining limits of distance between which there was a certain probability that the source occurred; for the run plotted in Fig. 7 the lines indicate the limits for a 50% probability, estimated from a knowledge of the

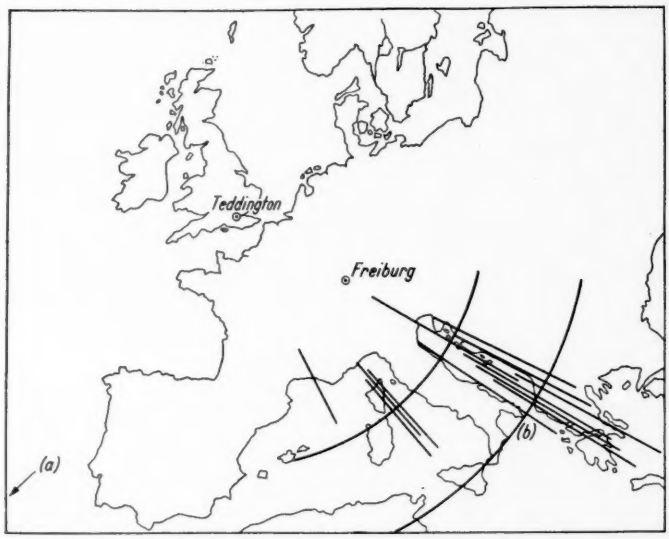


Fig. 7. Plot showing comparison between h.f. measurements and C.R.D.F. fixes at long range. 1500 hrs on 18 December 1950. Frequencies 12·05 and 14·15 Mc/s.
Note: The record for the period in which (a) and (b) occurred is shown in Fig. 3.

variability of the direction finding errors. The lines altogether extend over a large range in distance and it is difficult to estimate the distance of the storm or to decide whether the storm was localised or extended over a wide area.

The zone boundaries derived from ionospheric data from Freiburg (Germany) are plotted in Fig. 7. Of the more easterly group all except one atmospheric were received on the lower frequency only, indicating that they emanated from the intermediate zone. Of the atmospherics originating near Corsica two were received on the lower frequency only and one on neither frequency; that plotted near the south coast of France was received on the lower frequency only and was probably at a greater distance than was suggested by the C.R.D.F. plot.

These results indicate how, with sufficiently accurate ionospheric data, observations at high frequencies can reduce the uncertainties in the plotting of the direction finding bearings. In practice observations would be made at a number of frequencies to define the cut-off frequency, and therefore the range, more precisely.

4.4. *Sources in the outer zone*

In the outer zone atmospherics should be received at both frequencies except insofar as they are absorbed in the ionosphere. It is to be expected that the higher

VOL. 4 1953

frequency would tend to show the larger amplitudes, on the average, and this was generally, though not always, true.

When the sources were just outside the intermediate zone the amplitudes on the two frequencies were usually approximately equal. Examples may be quoted from the runs plotted in Figs. 5(*a*) and 6. In Fig. 5(*a*) the ratio of the amplitudes on the upper and lower frequencies were 0·37, 0·80, 0·85, 0·88 and 0·91 for the five flashes over the North Sea received on both frequencies. In Fig. 6 traces from (a), (b), (c) and (d) had ratios of 1·28, 0·56, 1·00 and 0·73 respectively.

An interesting set of results was obtained at 1400 on 18 October 1950. At frequencies of 12·05 and 14·15 Mc/s eighteen atmospherics recorded on both frequencies corresponded to C.R.D.F. fixes and were plotted in the outer zone which had an inner limit of 1000 km. They fell into three groups and the amplitude ratios (upper/lower frequencies) are listed in Table 1.

It will be seen that the amplitude ratios were quite different for the three storm centres although they were at similar ranges. These differences were probably due to the variations in the reflecting properties of the ionosphere which give rise to fading of signals.

Table 1

Position of storm		*No. of atmospherics*	*Amplitude ratio*	
Distance (km)	*Bearing (deg.)*		*Range of values*	*Average*
1700	217	10	1·5– 6·4	3·0
1850	143	5*	6·1–10·2	7·7
1540	164	3	9·2–15·0	11·6

The maximum range at which identification of corresponding atmospherics by C.R.D.F. and at high frequencies can be achieved would be expected to depend on the severity of man-made interference, on the intensity and frequency of atmospherics, and on the relation between the operating and cut-off frequencies. Conditions have varied considerably and it is not yet possible to quote a reliable estimate of maximum range. Four runs between 1100 and 1500 hours on different days showed a total of seven storm centres located by C.R.D.F. at ranges between 1600 and 2600 km. Three storm centres were in Eastern Europe, two in Spain (one plotted in Fig. 5) and two in the Atlantic. Of 44 atmospherics recorded by C.R.D.F. from these storms, only ten showed a possible trace at high frequencies, and only one could be regarded as satisfactorily identified. On the other hand the run on 18 October 1950 quoted earlier in this section showed 18 good identifications, with storm centres between 1500 and 1800 km, and there were possible traces at h.f. from a storm centre at 3000 km (over the Atlantic). The interference on this run appeared to be no less than at other times and similar frequencies were used.

* An additional atmospheric with a ratio of 6·0 was probably in this group but the identification was not definite.

The evidence regarding the limiting range for identification of flashes must therefore be regarded as inconclusive; identification may sometimes be good at 2000 km but is at other times poor at much shorter ranges. The use of directional aerials at high frequencies, selected to provide good reception from specific storm centres, would no doubt greatly improve identification.

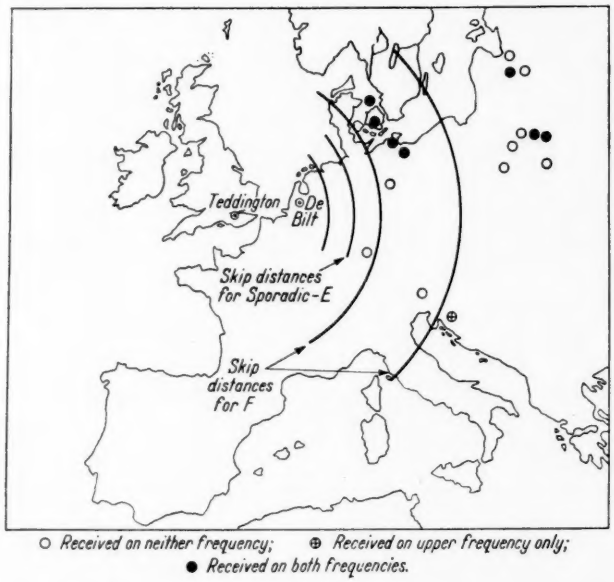

Teddington
De Bilt
Skip distances for Sporadic-E
Skip distances for F
○ Received on neither frequency; ⊕ Received on upper frequency only;
● Received on both frequencies.

Fig. 8. The influence of sporadic-*E* ionisation. 1500 hrs on 14 December 1950. Frequencies 16·58 and 19·82 Mc/s.
Note: The record for the period in which flashes (a), (b), (c) and (d) occurred is shown in Fig. 2.

4.5. *The influence of sporadic-E ionisation*

In all the results discussed so far, propagation by way of the *F*-layer was predominant, but on a few occasions sporadic-*E* ionisation has had some influence. An example is plotted in Fig. 8 in which a well-defined storm centre located by C.R.D.F. was plotted well inside the *F*-layer skip distance. Sporadic-*E* ionisation was reported at Poitiers (France) and Domont (France) however, and reception on the lower frequency only could be explained by taking this into account. According to Freiburg ionospheric data reception from the more easterly storm centre was controlled by the *F*-layer.

Another record plotted in Fig. 9 shows the possible influence of sporadic-*E* ionisation. The nearer group of points showed a curious mixture of flashes received on neither frequency and others received on both. According to De Bilt (Holland) ionospheric data energy should not have been received on the upper frequency by reflection at the *F*-layer, but due to sporadic-*E* ionisation both frequencies should have been received. The absence of received energy from some sources was probably due to the well-known non-uniformity of the sporadic-*E* ionisation.

VOL. 4 1953

4.6. *Anomalous results*

Most of the results showed general agreement with the ionospheric data, but a few records could be explained only by assuming abnormally large errors either in the C.R.D.F. bearings or in the interpretation of the ionospheric data. In one such run (1500 in 8 December 1950) the skip distances were 1330 km and 1890 km and two flashes received only on the lower frequency were plotted at ranges of 700 km

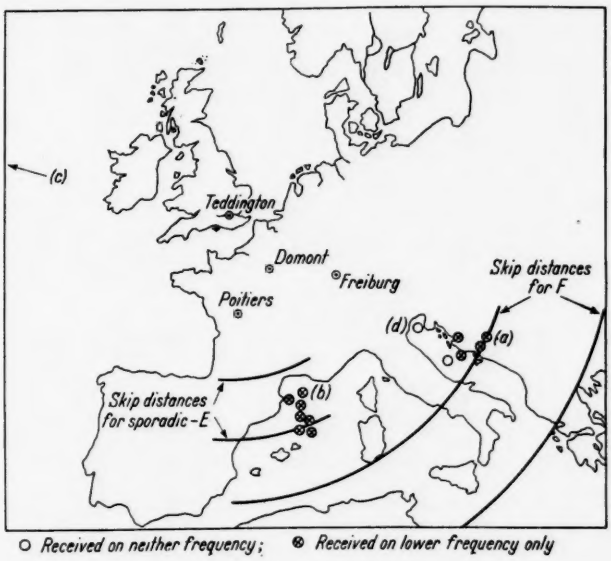


Fig. 9. Plot showing possible effects of sporadic-*E* ionisation. 1100 hrs on 23 June 1950. Frequencies 10·03 and 12·19 Mc/s.

and 900 km. In the absence of other evidence the two fixes would have been considered to be in error by not more than 100 km. On the same run but a different bearing two atmospherics received only on the lower frequency were plotted in the intermediate zone and therefore agreed with the ionospheric data.

On some occasions atmospherics just outside the skip zones have been received strongly on the higher frequency and much less strongly on the lower frequency. As the difference in absorption should be small these results are considered to be anomalous.

It would be unwise to attach too great importance to these results. They are considered to be anomalous at this stage because the identifications appeared to be good and the disagreement between the fix and the propagation data much greater than could be attributed to errors of usual magnitude. However identification of traces can never be absolutely certain, and large errors in C.R.D.F. fixes do sometimes occur, so the occasional apparent anomalies may be no more than an indication of the present state of the technique.

5. Discussion of the Technique as an Aid to Thunderstorm Location

The results of this preliminary investigation have shown that while there are practical difficulties in interpreting simultaneous high frequency observations and low frequency directional measurements, the reception of atmospherics is in general agreement with theoretical deductions based on the known ionospheric

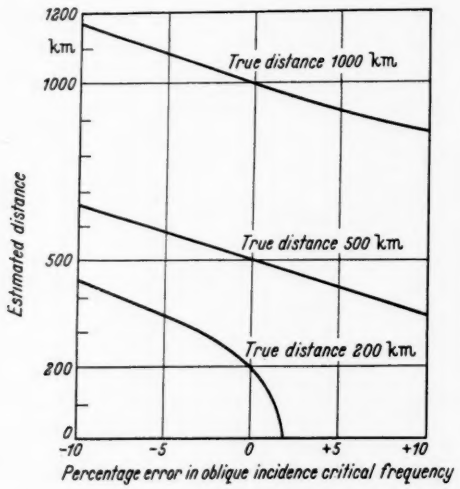


Fig. 10. Errors in distance estimate due to errors in estimating the oblique incidence critical frequency.

conditions. Let us now suppose that the practical difficulties have been overcome and that simultaneous recordings of atmospherics are available at a sufficient number of frequencies for the cut-off frequency to be determined accurately. At times near to sunrise and sunset this frequency may not define the skip distance accurately for an east-west path because of the rapidly changing ionospheric conditions. It will be assumed that at other times of the day and night the ionospheric conditions are nearly uniform between the storm and the observing point. The accuracy of the derivation of the distance of the source then depends on the accuracy to which the ionospheric conditions at the mid-point of the path can be estimated. Errors arise because the conditions must be derived by interpolation between ionospheric observatories, because oblique incidence conditions are derived from vertical incidence measurements, and because there are always some variations in ionospheric conditions as a function of time.

The effect of errors in estimating the oblique incidence critical frequency is shown in Fig. 10. For this illustration it has been assumed that storms exist at ranges of 200, 500 and 1000 km from the observing point and that the oblique incidence critical frequency has been measured for each. The errors in distance estimation resulting from errors in the calculated critical frequencies are plotted in the figure, values for the height and thickness of the reflecting layer being

VOL. 4 1953

assumed correct. At the shortest range (200 km) a small overestimation of the calculated critical frequency leads to very large errors, but at distances of 500 km upwards an error of $\pm$ 5 per cent in critical frequency leads to an error of only about $\pm$ 80 km in distance. The percentage accuracy therefore improves with increase in distance. On the other hand the errors involved in interpolating ionospheric data between stations tend to increase with distance, but on balance the technique is more applicable to the longer ranges and is therefore complementary to direction finding techniques with relatively short baselines.

In the experiments described the results of ionospheric soundings made at the time of the run were used. For many research purposes in which the positions of the storms need not be known until some time after the observations, the use of soundings is permissible. For routine location of thunderstorms, however, where the results are required quickly (for weather forecasting for example) it would be necessary to use predicted ionospheric data, based perhaps on vertical incidence soundings taken an hour or two earlier. Examination of typical hourly records of vertical incidence critical frequency at Slough shows that hour-to-hour changes are usually within 10%, and by observing trends the values might be predicted to a much better accuracy. If the overall accuracy of the estimated oblique incidence critical frequency can be reduced to a few per cent, the accuracy of distance estimation becomes comparable with or better than that obtained by C.R.D.F. at ranges of about 1000 km. Considering only the limitations set by the accuracy of ionospheric data, therefore, it appears that the technique merits further investigation.

The use of back-scatter measurements (Villard and Peterson, 1952) for obtaining information about the ionosphere should also be considered. This technique involves the use of a pulse transmitter radiating energy in a narrow beam whose direction can, preferably, be varied continuously over a wide azimuth range. The pulses are returned to the observing point by reflection of sky-waves at the ground at distances greater than the skip distance. Observations of the minimum time delay of the echoes indicates the skip distance for the particular frequency used. This type of measurement therefore yields ionospheric information directly in the form required for the atmospherics application.

Turning now to the practical difficulties of obtaining a measurement of oblique incidence critical frequency, the photographic technique is no doubt the best method for comparing the reception on different frequencies. The analysis of film records is slow, however, and more rapid methods should be considered. Direction finding observations are made at present by reading the bearings of traces on a cathode ray tube with a persistent screen and it may be possible to display the flashes received at high frequencies in a similar way and to compare their amplitudes visually. The method, however, would be dependent on more satisfactory identification of the atmospherics and on the reduction of interference.

6. Conclusions

The preliminary investigation into the possibilities of a high-frequency ranging technique has produced encouraging results and it is considered that a further and more elaborate investigation would be justified. It appears that the technique

would not always be satisfactory at times near sunrise and sunset but that useful accuracy of distance measurement can be achieved at other times of the day. Further investigation would involve measurements on a site free from man-made interference, the use of directional aerials to simplify identification, and the use of ionospheric data derived from back-scatter experiments. To eliminate the time delay inherent in the method of identification used so far, the C.R.D.F. bearings could be recorded on film, with an identification signal superimposed on both the low frequency and high frequency records. The possibilities of a display involving visual observation of the cathode ray tubes should also be explored.

A more complete investigation of short-term variations in ionospheric conditions is also required, including an estimate of the accuracy to which conditions can be predicted an hour or two in advance.

Acknowledgments—The work described above was carried out as part of the programme of the Radio Research Board. This paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research. The experiments were carried out by Mr. B. J. BYRNE.

The recording of bearings of atmospherics by means of the Meteorological Office network of direction finders was an essential part of the investigation; the assistance provided by the Director of the Meteorological Office and his staff is gratefully acknowledged.

VOL.
4
1953

REFERENCES

CLARKE, C., and MORTIMER, D. E. 1951 Automatic Atmospherics Waveform Recorder. Wireless Engineer **28**, 359–370

VILLARD, O. G., and PETERSON, A. M. 1952 Instantaneous Prediction of Radio Transmission Paths. Q.S.T. **36**, 11–20

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 141 to 147. Pergamon Press Ltd., London

Determination of the location of the ionospheric current system responsible for geomagnetic effects of solar flares*

A. P. MITRA** and R. E. JONES

Ionosphere Research Laboratory, The Pennsylvania State College

(*Received* 25 *June*, 1953)

ABSTRACT

A new method for locating the ionospheric current system associated with solar flares is presented. The method depends on the time of maximum intensity of a geomagnetic flare effect and the enhancement of electron density at the relevant level.

It is found that the flare current system is located either at 100–130 km (if one assumes that the flare effect is essentially one of enhancement of the *Sq*-current) or at 60 km (if it is an independent system).

INTRODUCTION

It is well known that during strong solar flares there are sometimes sudden changes in the three components H, D and V of the terrestrial magnetic field. These are known as geomagnetic flare effects and sometimes as "crochets." MCNISH (1937a and 1937b) has made a survey of the geomagnetic effects, recorded over a whole hemisphere, produced by three major flares which occurred in 1936 and concluded that the geomagnetic effect is an augmentation of the S-field of the sunlit hemisphere. Such augmentation presumably occurs because of an enhancement of ionization at certain levels, resulting in increased strength of the current system. On this hypothesis the geomagnetic flare effects and the various sudden ionospheric disturbances (S.I.D.'s) are related with each other. Subsequent investigations by a number of authors such as IMAMITI (1938 and 1943), NAGATA (1950, 1952), NEWTON (1948) and MCINTOSH (1951) have generally confirmed the above view-point. Further, it is believed (MCNISH, 1937a and 1937b; IMAMITI, 1938 and 1943; NAGATA, 1950 and 1952; NEWTON, 1948) that the excess current is in the same direction as the normal diurnal variation current.

Whether this excess current is an independent system or is an enhancement of the normal *Sq*-current system is not definitely known. However, the consensus is that the excess current is a different current system located presumably in the D-layer the ionization of which is known to increase during an S.I.D.† (CHAPMAN and BARTELS, 1940), whereas the normal *Sq*-current is located in the E-layer around 100 km level (MAPLE, BOWEN, and SINGER, 1952; BAKER and MARTYN, 1952).

In the present note a new method is presented for locating the flare current system. The method depends on the difference between the maximum intensity of a geomagnetic flare effect and of that of the flare causing it.

---

* The research reported in this note has been sponsored by the Geophysical Research Division of the Air Force Cambridge Research Center, Air Research and Development Command under Contract AF19(122)-44.

** On leave from the Council of Scientific and Industrial Research, Government of India and the Institute of Radiophysics and Electronics, University of Calcutta.

† E. V. APPLETON and W. R. PIGGOT, private communication.

2. METHOD

There is a "relaxation" time, τ, associated with the production and disappearance mechanisms of ionization for any level in the ionosphere. This time is defined here as the time delay between the maximum of the crochet and the corresponding flare. This time lag is primarily a recombination effect and is given approximately by MITRA and JONES (1953)

$$\tau(h) = \frac{1}{2\alpha(h)N(h)} \tag{1}$$

where α and N represent the recombination coefficient and electron density at the height h. Using the subscripts n and f for conditions during normal and flare times, respectively, and the notation $\bar{h}$ for the height of the current system, we obtain

$$\frac{\tau_n(\bar{h})}{\tau_f(\bar{h})} = \frac{N_f(\bar{h})}{N_n(\bar{h})} \tag{2}$$

Since $N_f(\bar{h}) > N_n(\bar{h})$, then $\tau_f(\bar{h}) < \tau_n(\bar{h})$; that is, during a flare there is a decrease in the value of the delay time associated with that level, the magnitude of the decrease depending on the enhancement of ionization following a flare. Hence, given $\tau_f(\bar{h})$, $\tau_n(\bar{h})$ and $\dfrac{N_f(\bar{h})}{N_n(\bar{h})}$, it is possible to estimate $\bar{h}$, the height of the current system producing the magnetic effects.

The value of τ_n associated with any height may be determined from the diurnal variations of such ionospheric data as ionization, height, and polarization. BRACEWELL and BAIN (1952) have reported an asymmetry of "D_α" phase heights corresponding to 25 ± 5 minutes. The probable height involved is about 70 km. BENNER (1951) has studied the noontime delay in the absorption maximum for 150 Kc/s and has given a value of about 60 minutes for the most probable delay time. The corresponding height is about 80 km. Analysis of the polarization tilt angle as given by NEARHOOF (1951) or KELSO *et al.* (1951) for 150 Kc/s gives a delay time of about 80 minutes. This value also refers to a height of about 80 km. JONES (1952) has given monthly average diurnal variations of the phase height of the main E-region reflection on 150 Kc/s. The noontime delay was found to be some 30 minutes, the corresponding height being about 90 km. Finally, the delay time, for a height of about 110 km, may be obtained from the routine critical frequency data for the E-region. MITRA and JONES (1953) have analyzed the f_0E data for a number of stations as reported in the CRPL-F Bulletins.* The average value obtained by them is about 4 minutes.

The above values (listed in Table 1) permit one to draw an approximate distribution of τ_n with height. Such a distribution is shown in Fig. 1. It may be mentioned here that two other values of τ_n are available for ionospheric heights. One is due to PIGGOT† (2000 seconds) and has been obtained from absorption measurements made at Slough for medium short waves. The other ($\approx$ 20 min.) has been

* Central Radio Propagation Laboratory, National Bureau of Standards, Bulletins CRPL-F 54 through F 90.

† W. R. PIGGOT, private communication.

VOL. 4 1953

obtained by MITRA and SHAIN (1953) from absorption measurements of cosmic noise at 18·3 Mc/s. It is possible, however, that in both these cases, appreciable contamination by E-absorption exists—a fact which makes assignment of proper heights to these values difficult. These results have therefore been excluded from Fig. 1.

The experimental τ_n-h curve in Fig. 1 covers only the height range from 70 km to 110 km. It is possible, however, to extend the curve on both sides by making

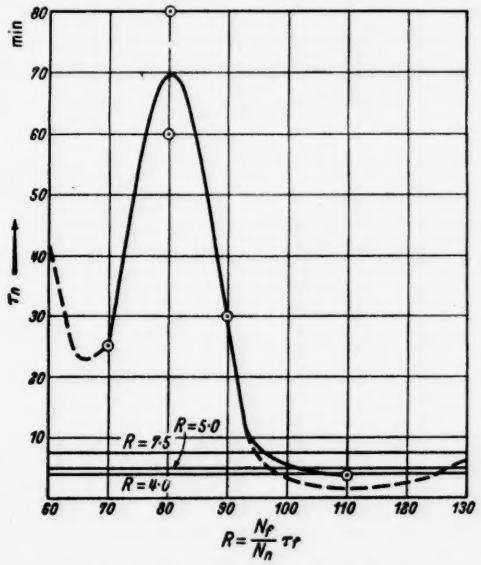


Fig. 1. Height variation of τ_n. Full line portion obtained from experimental data and dotted line from theoretical considerations.

some reasonable assumptions regarding the electron distribution in the regions involved. In D-region τ_n was calculated from the recombination coefficient model given by MITRA and JONES (1953) and the D-region electron density model of

Table 1. *Relaxation time as a function of height*

	Authors	*Method*	τ_n (min)	*Height* (km)
1.	BRACEWELL and BAIN	"D_α" phase heights	25 ± 5	70
2.	BENNER	150 kc absorption	60	80
3.	Calculated from NEARHOOF	150 kc polarization	80	80
4.	Calculated from JONES	150 kc phase heights	30	90
5.	MITRA and JONES	f_0E	4	110

NERTNEY (1953) (Fig. 2). On the E-region side the calculation was made for a CHAPMAN model with $\chi = 0°$ and $\chi = 45°$. These extensions are shown in dotted lines in Fig. 1.

As regards $\tau_f(\bar{h})$, this quantity may be estimated from the difference in the times of maxima of the geomagnetic flare effect and the corresponding flare. In general, the difference is measureable. ELLISON (1950) has recorded the times of beginning, maximum, and end of six large geomagnetic flare effects, of which

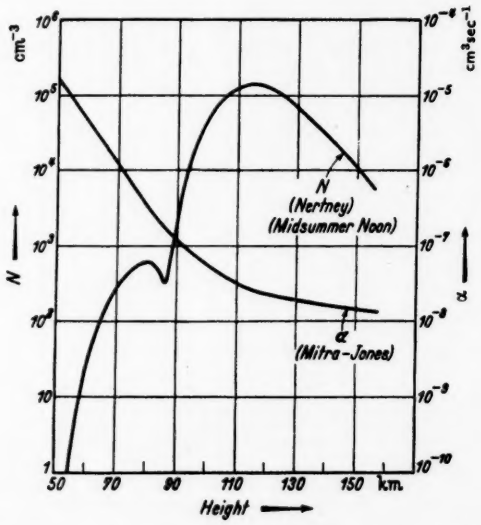


Fig. 2. Height variation of recombination coefficient (α) and electron density (N).

five were produced by flares of classes 3 and 3+ and one by a flare of class 2. On the average the maximum of the geomagnetic effect occurred 2·5 minutes after the occurrence of the peak line-width of the flare. The standard error is approximately 1·5 minutes. A further corroboration of this value is obtained from the list of geomagnetic solar flares given by MCINTOSH (1951) for Lerwick and Eskdalemuir over the period 1936–49. An average delay time, obtained from his list for cases where the time of maximum flare intensity was known, has the value of about 3 minutes; while Fig. 3 of his paper, in which the times of growth and decay of the flares and various flare effects, including the geomagnetic effects, have been given, shows a delay time of about 2·5 minutes. All the evidence is, therefore, for a time delay, at the time of the flare, of about 2·5 minutes for the level at which the current producing the geomagnetic effect flows.*

Consider first the case in which the flare current is at the same level as normal Sq-current. A reliable estimate of $\dfrac{N_f(\bar{h})}{N_n(\bar{h})}$ may be formed from the magnitude of

* Because of possible induction effect in the ionosphere, the observed delay time has to be considered as an upper limit for delay caused by recombination.

VOL. 4 1953

the terrestrial magnetic variations occurring at that time. McNish (1937a) has presented, in a number of diagrams, the direction and amplification of the overhead current system of flaretime effects over a number of stations in the northern hemisphere for three major flares in 1936. In these cases the electric currents in the ionosphere appear to be doubled suggesting, therefore, a value of about 2 for the ratio $\frac{N_f(\bar{h})}{N_n(\bar{h})}$. A survey of the list presented by McIntosh also gives similar values. For intense flares the amplification may be larger, but a value of 3 seems a safe upper limit on the basis of the available data.

We are now in a position to locate the current system. In Fig. 1 lines for $\frac{N_f(\bar{h})}{N_n(\bar{h})}\tau_f(\bar{h}) = 4{\cdot}0,\ 5{\cdot}0,\ 7{\cdot}5$ (corresponding to $\tau_f(\bar{h}) = 2{\cdot}5$ minutes and $\frac{N_f(\bar{h})}{N_n(\bar{h})} = 1{\cdot}6,\ 2{\cdot}0,\ 3{\cdot}0$) have been drawn. The intersections of these lines with the curve of τ_n will identify the height of the system. It may be seen that the intersection occurs in the neighbourhood of 100–120 km. Even for the upper limit of $N_f/N_n = 3$, which is not likely to be usually reached, and for the lower curve, the range of intersection is within 100–130 km.

In order that there be a second region of intersection in the *D*-layer, the values of τ_n there would have to be much lower than calculations indicate. For this to happen, the electron density in the relevant portions of the *D*-region would have to be more than three times as great as in the model chosen. However, even if this should occur, the lower current system is very unlikely to produce sufficient geomagnetic effects, in view of the low conductivity of the region as indicated by Bates and Massey (1951). One may therefore conclude (on the hypothesis that the flare effect current is an enhancement of the normal *Sq*-current) that the current system causing the geomagnetic flare effects (as well as the *Sq*-current system normally present) is localized in the region 100–120 km.

It is interesting to compare this result with those obtained by other methods. From rocket measurements at the magnetic equator, Maple, Bowen and Singer (1952) have obtained direct evidence of a current system of considerable strength near the 100 km level but none of appreciable magnitude below. In a more recent work Baker and Martyn (1952) have attempted to explain the observed ionospheric conductivity by taking account of vertical currents prohibited by polarization, and have shown that the current will be located at 100 ± 10 km for regions near the geomagnetic equator and between 120–150 km in higher latitudes (depending on the ionization distribution and tidal velocities at these levels). The agreement of these results with ours is considered significant.

3. Case of an Independent Flare Effect Current

If the current system responsible for the flare effect *is* an independent system, then the arguments presented above will no longer hold. McIntosh (1951) has recently presented some evidence in favour of such a hypothesis. In a survey of the geomagnetic flare effects at Lerwick and Eskdalemuir over the period 1936–49, McIntosh found that, although the *H* and *D* components point strongly to the conclusion that the geomagnetic flare and *Sq*-effects are produced by the same

direction of ionospheric current, the V-components of the effects and of the normal variations were considerably different and indicated a substantial difference between the space distribution of the currents in the two cases. If we accept this evidence, then the observed amplification of the Sq-current system will no longer represent the enhancement of electron density at *one* particular level, but will be an effect contributed by two separate regions located at two different heights. Under these conditions it would be difficult to estimate $\frac{N_f(\bar{h})}{N_n(\bar{h})}$ at the level of the extra current system from the observed geomagnetic changes. Equation (2), however, still holds and, provided the right hand side can be determined, may still be used.

Now, there are other ways of estimating the enhancement of electron density. One of these is the measurement of ionospheric absorption. APPLETON and PIGGOT* have made some measurements regarding the excess attenuation at the time of flares for medium short waves. They have observed a total increase of ten times the normal value in their absorption data for S.I.D.'s lasting between one-half and one hour. SHAIN and MITRA (unpublished) have made a number of observations on the increase in absorption, during flares, of cosmic noise at 18·3 Mc/s. For S.I.D.'s lasting about one hour the excess attenuation in their observation was some four times the normal value. Since most of this excess absorption may probably be attributed to D-layer, it appears that an enhancement of ionization by a factor of about five times or more may occur for intense flares producing the geomagnetic effects.

An enhancement of this magnitude is capable of approximating the observed delay time ($\approx$ 2·5 minutes) at the height range 60–70 km. That low values of τ_f are, in fact, encountered at heights around 60 km may be observed in Table 2. This table has been prepared from the curves given by ELLISON (and also from his list) for various sudden phase anomalies and sudden enhancement of atmospherics accompanying several major flares.

Table 2. Average time delay (observed) from various S.I.D. phenomena for some major flares in 1949 (ELLISON)

Effect	*Class* 3 *and* 3+		*Class* 1 *and* 2	
	τ_f	*Appr. height*	τ_f	*Appr. height*
S.P.A.	3 min	54 km	7 min	62 km
S.E.A.	6		9	
G.B.R. (Ed†)	5		8	

* Private communication.

† These results are obtained from records of long range fading due to ground wave-sky wave interference. The transmitter is G.B.R. at Rugby (16 kc) and the receiver is at Edinburgh, 420 km away.

VOL. 4 1953

Although the values given in Table 2 are somewhat larger than the required 2·5 minutes, little significance should be attached to this small discrepancy which is not much larger than the possible errors of measurement.

Thus it appears that an extra current system located at about 60 km may have the correct delay time for flare effects. Whether such a current system can have adequate conductivity to give rise to the large geomagnetic effects is another question, and calculations have to be made, perhaps along the lines of BAKER and MARTYN, before any conclusions may be attempted.

In view of the above, the following conclusions may be made: (i) If one assumes that the flare effect is essentially one of an enhancement of the *Sq*-current system, then the location of this system, according to the present results, will be between 100–130 km. (ii) In the event that the flare current is an independent system, the location of this current will be around 60 km, while the *Sq*-current will be at about 100–130 km.

REFERENCES

BAKER, W. G., and MARTYN, D. F.	1952	Nature **170,** 1090
BATES, D. R., and MASSEY, H. S. W.	1951	J. Atmosph. Terr. Phys. **2,** 1
BENNER, A. H.	1951	Technical Report No. 18, Ionosphere Research Laboratory, State College, Pa., Jan. 30, 1952; or Proc. Inst. Rad. Eng. **39,** 186–190
BRACEWELL, R. N., and BAIN, W. C.	1952	J. Atmosph. Terr. Phys. **2,** 216–225
CHAPMAN, S., and BARTELS, J.	1940	Geomagnetism, Clarendon Press, Oxford, p. 791.
ELLISON, M. A.	1950	Royal Observatory **1,** 53–72
IMAMITI, S.	1938	Mem. Kakioka Mag. Obs. **1,** 13
	1943	**5,** 45
JONES, R. E.	1952	Scientific Report No. 41, Ionosphere Research Laboratory, State College, Pa., August 15
KELSO, J. M., NEARHOOF, H. J., NERTNEY, R. J., and WAYNICK, A. H.	1951	Ann. Géophys. **7,** 215–244
MAPLE, E., BOWEN, W. A., and SINGER, S. F.	1950	J. Geophys. Res. **55,** 115
	1951	**56,** 265
	1952	Nature **170,** 1093
MCINTOSH, D. H.	1951	J. Atmosph. Terr. Phys. **1,** 315
MCNISH, A. G.	1937a	Terr. Mag. **42,** 109
	1937b	J. Appl. Phys. **8,** 718–731
MITRA, A. P., and JONES, R. E.	1953	Scientific Report No. 44, Ionosphere Research Laboratory, State College, Pa., March 25
MITRA, A. P., and SHAIN, C. A.	1953	J. Atmosph. Terr. Phys. (in press)
NAGATA, T.	1950	Rep. Ion. Res. Jap. **4,** 155
	1952	J. Geophys. Res. **57,** 1
NEARHOOF, H. J.	1951	Technical Report No. 25, Ionosphere Research Laboratory, State College, Pa., Aug. 20
NERTNEY, R. J.	1953	J. Atmosph. Terr. Phys. **3,** 92
NEWTON, H. W.	1948	Mon. Not. R. Astron. Soc., Geophys. Suppt. **5,** 200–215

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 148 to 162. Pergamon Press Ltd., London

Bailey's theory of sunspot noise

J. W. DUNGEY*

School of Physics, University of Sydney, N.S.W.

(*Received* 9 *July* 1953)

ABSTRACT

The exact formulation of BAILEY's theory for a non-uniform ionised gas is described and the boundary conditions are explained. A direct calculation is made of the rate of exchange of energy between the motion of the electrons and the wave. The intensity of the radiation arising from a transient disturbance is calculated and it is found that the intensity can become very large when the wave gains energy from the electrons. The general nature of the electron flow and the order of magnitude of the electron velocity needed for amplification to occur are discussed.

1. INTRODUCTION

After developing the theory of radio waves in an ionised gas including the effect of static electric and magnetic fields Professor V. A. BAILEY (1948) proposed a theory which accounted for the intense radio noise emitted by sunspots as due to amplification by the medium (BAILEY, 1950). This theory has been criticised by TWISS (1950) who states that BAILEY'S growing waves must be regarded as reflected waves. TWISS bases this statement on a study of the transient response of a uniform medium, but it may be investigated more directly by considering the effect of the boundary condition, namely that

> far from the source the waves must
> be pure outgoing waves. (A)

The actual medium is of course non-uniform and BAILEY (1950) has considered a model in which the medium is represented as a series of uniform slabs, and has given the continuity conditions for the discontinuous boundaries between the slabs. For the model with just one discontinuity, it is found that when there is only an outgoing wave on the side away from the sunspot, no amplification is possible. For the model with two discontinuities amplification is possible, but only under special conditions, for instance the distance between the discontinuities must be within a very small range of values. The disagreement between the results obtained for these two models however shows that TWISS's arguments from the first model are not conclusive, and it was therefore decided to work out the theory of a continuously varying medium, which is the subject of this paper. Ideally one needs to work with a non-steady motion, varying in time as well as space, but this involves great mathematical difficulties even in very special cases. In this paper a rather general discussion of the case of steady motion is given, and this gives an indication of what may be expected to happen in the case of non-steady motion.

* Now at Dept. of Electrical Engineering, Pennsylvania State College.

VOL. 4 1953

For models satisfying the special condition of zero vorticity BUNEMAN (1951) has shown that spontaneous waves cannot be radiated. The present work is more general, however, so that this condition does not in general hold, and the conclusions obtained in fact show that BUNEMAN'S result cannot be extended to the general case.

Before describing the model to be used, its relation to the phenomena of solar noise should be explained. The explanation of the radiation from the quiet sun as thermal noise is generally accepted, but the various types of storm, or "enhanced radiation," have received no generally accepted explanation. The suggestion that enhanced radiation is of thermal origin, or that it is emitted by cosmic rays which are generated by solar flares, is unlikely to account for all the noise storms, and it will be supposed in this paper that the temperature of the gas is much too low to account for the strength of the radiation. Another possibility has been studied by JAEGER and WESTFOLD (1949), who have shown that the spectrum and development in time of isolated bursts can be accounted for if they are caused by sudden disturbances at suitable heights in the solar atmosphere, but no estimate of the magnitude of such disturbances to be expected in the corona has been given. The present work will not remove this difficulty, but will show how the radiation may be amplified by absorbing energy from the motion of the electrons in the solar atmosphere as suggested by BAILEY, so that the strength of the initial disturbance required to explain the observed intensity may be much smaller than it would be otherwise. The general nature and magnitude of the electron motion required for amplification is deduced; whether it actually occurs is not known, but it is not impossible, at any rate in a non-steady state.

Recently TWISS (1952) has given a qualitative account of a similar process using the terminology of radio circuits. In the present paper the fundamental equations of the electro-magneto-ionic theory are used in conjunction with the boundary condition (A) to determine what the wave looks like in the solar atmosphere, using a fairly general model. The method used is a natural development of BAILEY'S, and the suggestion that a radio wave may absorb energy from the electron motion is also his, but the amplification considered in this paper occurs in regions where the refractive index is real. Amplification can also occur in overdense regions, but this is not further discussed here as it requires a considerable extension of the present investigation.

The model for the solar atmosphere is restricted from the outset as follows. The electron density N and velocity $\boldsymbol{U}$ and the magnetic field $\boldsymbol{H}$ are independent of time. With the z-axis pointing out of the sun, N varies with z only, and $\boldsymbol{H}$ is uniform and parallel to the z-axis. The simplification resulting from this restriction of $\boldsymbol{H}$ is thought to outweigh the loss of generality because it is found that amplification can occur without any magnetic field. With $\boldsymbol{U}$ unrestricted an expression is obtained for the energy exchange between the electron motion and the wave. The effect on the component of a given frequency arising from a sudden disturbance is then considered, and it is seen that this component will be strongly amplified if a certain resonance condition is fulfilled. The possibility of achieving this condition is then discussed in greater detail using a more restricted model in which $\boldsymbol{U}$ is axially symmetric about the z-axis.

2. THE FUNDAMENTAL EQUATIONS

The fundamental equations are obtained in the same way as BAILEY'S, but the spatial variation of the wave-components cannot now be expressed by means of a constant wave-number. The wave is described by components $\boldsymbol{e}$ and $\boldsymbol{h}$ of the electric and magnetic fields and n and $\boldsymbol{u}$ of the electron density and velocity, all of which are taken to vary with time as $e^{i\omega t}$; the wave field is the real part of these components so that, for example, if N is the total electron density

$$\mathsf{N} = N + \tfrac{1}{2}(n + n^*)$$
$$\frac{\partial \mathsf{N}}{\partial t} = \tfrac{1}{2}i\omega(n - n^*)$$

The effect of the positive ions is much smaller than that of the electrons because of their greater mass, and is neglected. Then from MAXWELL'S equations we obtain by picking out the terms proportional to $e^{i\omega t}$

$$i\omega \boldsymbol{h} = -c \operatorname{curl} \boldsymbol{e} \tag{1}$$

$$c \operatorname{curl} \boldsymbol{h} - i\omega \boldsymbol{e} = 4\pi q(N\boldsymbol{u} + n\boldsymbol{U}) \tag{2}$$

where q is the electronic charge.

The term $n\boldsymbol{u}$ omitted from (2), which is proportional to $e^{2i\omega t}$ and would therefore set up harmonics, is regarded as negligible, and similar non-linear terms are neglected in the other fundamental equations which are the continuity equation

$$i\omega n = -\operatorname{div}(N\boldsymbol{u}) - \operatorname{div}(n\boldsymbol{U}) \tag{3}$$

and the equation of transfer of momentum

$$(i\omega + \nu)\boldsymbol{u} + (\boldsymbol{U}\cdot\nabla)\boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{U} = \frac{q}{m}(\boldsymbol{e} + \boldsymbol{U}\wedge\boldsymbol{h}/c) - \boldsymbol{u}\wedge\boldsymbol{\Omega} \tag{4}$$

where $\boldsymbol{\Omega} = -q\boldsymbol{H}/mc$ and the term $\nu\boldsymbol{u}$ represents the effect of collisions; the "collision frequency" ν depends on the temperature and density and will be taken to depend on z only.

The general equations have no simple exact solutions, but the radiation from the sun is of course transverse and often circularly polarised, so that solutions which approximate to circularly polarised waves travelling in the z-direction are of interest. With no electron drift velocity these would be exact solutions and it is convenient to introduce the notation

$$\left.\begin{aligned} e_\pm &= (e_x \pm ike_y)/\sqrt{2} \\ h_\pm &= (h_x \pm ikh_y)/\sqrt{2} \\ u_\pm &= (u_x \pm iku_y)/\sqrt{2} \end{aligned}\right\} \tag{5}$$

where $k = -\Omega_z/\Omega$

and $\Omega = |\boldsymbol{\Omega}|$.

Remembering that q is negative, it is found that the components with subscripts $+$ and $-$ refer to ordinary and extraordinary circularly polarised waves

VOL. 4 1953

respectively.* Then, when $\boldsymbol{U} = 0$, (4) reduces to

$$(i(\omega \pm \Omega) + \nu)u_{\pm} = \frac{q}{m} e_{\pm} \tag{6}$$

Most of the work in this paper refers to ordinary circularly polarised waves. Also, it will be shown in section 6 that ν is small compared with ω, so that in a first approximation (6) yields

$$u_{+} \approx -\frac{iq}{m(\omega + \Omega)} e_{+} \tag{7}$$

3. The Exchange of Energy Between the Electron Drift Motion and the Wave

Bailey (1950) discussed the exchange of energy between the electron drift motion and the wave by means of the Maxwellian pressure. The discussion in this section is similar but more general, and it is now useful to introduce the mean Poynting vector $\boldsymbol{P}$ belonging to the wave.† Since it is proportional to the vector product of the electric and magnetic fields, the time average is obtained from those products which are independent of time, thus

$$\boldsymbol{P} = \frac{c}{16\pi} (\boldsymbol{e} \wedge \boldsymbol{h}^* + \boldsymbol{e}^* \wedge \boldsymbol{h}) \tag{8}$$

(1) and (2) and their complex conjugates then yield

$$\operatorname{div} \boldsymbol{P} = -\frac{q}{4} [N(\boldsymbol{u} \cdot \boldsymbol{e}^* + \boldsymbol{u}^* \cdot \boldsymbol{e}) + \boldsymbol{U} \cdot (n\boldsymbol{e}^* + n^*\boldsymbol{e})] \tag{9}$$

which shows that div $\boldsymbol{P}$ is equal to the mean rate of work done by the electrons against the electric field; this is the fundamental property of the Poynting vector.

On taking the scalar product of (4) with $\boldsymbol{u}^*$ and adding the complex conjugate we obtain

$$\begin{aligned} 2\nu\boldsymbol{u} \cdot \boldsymbol{u}^* + (\boldsymbol{U} \cdot \nabla)(\boldsymbol{u} \cdot \boldsymbol{u}^*) + \boldsymbol{u}^* \cdot (\boldsymbol{u} \cdot \nabla)\boldsymbol{U} + \boldsymbol{u} \cdot (\boldsymbol{u}^* \cdot \nabla)\boldsymbol{U} \\ = \frac{q}{m} (\boldsymbol{u}^* \cdot \boldsymbol{e} + \boldsymbol{u} \cdot \boldsymbol{e}^* + \boldsymbol{U} \cdot (\boldsymbol{h} \wedge \boldsymbol{u}^* + \boldsymbol{h}^* \wedge \boldsymbol{u})/c) \end{aligned} \tag{10}$$

If $\boldsymbol{U} = 0$, (9) and (10) give

$$\operatorname{div} \boldsymbol{P} = -\tfrac{1}{2}\nu m N \boldsymbol{u} \cdot \boldsymbol{u}^*$$

which confirms the well-known fact that the only energy exchange between the wave and the electron gas is that due to collisions.

If $\boldsymbol{U} \neq 0$, there is an additional contribution to div $\boldsymbol{P}$ which is of first order in $\boldsymbol{U}$. To obtain the rate of exchange of energy between the wave and the electron gas,

* This agrees with Bailey's (1950) definition of E_1 and E_2 waves which are extraordinary and ordinary respectively.

† The plain symbol $\boldsymbol{P}$ is used for the *mean* Poynting vector.

however, we must now include the flow of wave kinetic energy (of the electrons); the mean rate of this flow is $\boldsymbol{F}$ defined by

$$\boldsymbol{F} = \frac{m}{4}\{N[(\boldsymbol{u}\cdot\boldsymbol{u}^*)\boldsymbol{U} + (\boldsymbol{U}\cdot\boldsymbol{u})\boldsymbol{u}^* + (\boldsymbol{U}\cdot\boldsymbol{u}^*)\boldsymbol{u}]$$

$$+ n[(\boldsymbol{U}\cdot\boldsymbol{u}^*)\boldsymbol{U} + \tfrac{1}{2}\boldsymbol{U}^2\boldsymbol{u}^*] + n^*[(\boldsymbol{U}\cdot\boldsymbol{u})\boldsymbol{U} + \tfrac{1}{2}\boldsymbol{U}^2\boldsymbol{u}]\} \qquad (11)$$

terms of fourth power in the wave amplitude being neglected. The rate at which energy is transferred from the electron gas to the wave per unit volume is then div $(\boldsymbol{P} + \boldsymbol{F})$, and this can be obtained from (9), (10) and (11). The terms containing N in (11) partially cancel the terms containing $\boldsymbol{U}$ on the left-hand side of (10), but the complete expression for div $(\boldsymbol{P} + \boldsymbol{F})$ is long and will not be written out here; an approximate expression for a special case is given by (12). The important conclusions are that the part involving $\boldsymbol{U}$ does not vanish identically and that it can be positive, if the electron drift motion has a suitable nature; this proves that it is possible in principle for the wave to gain energy from the electron drift motion. Unfortunately it is not easy to picture the physical mechanism by which this exchange occurs, but it is hoped that the following discussion may shed some light on this question.

The exchange of energy is due to work done on or by the electrons, and this depends on the component of $\boldsymbol{u}$ which is in phase (or antiphase) with $\boldsymbol{e}$; this can be obtained from equation (4), which shows that, if ν, $\boldsymbol{U}$ and $\boldsymbol{\Omega}$ all vanish $\boldsymbol{u}$ is in quadrature with $\boldsymbol{e}$. There are two separate effects caused by the drift velocity, and they are represented by the terms containing $\boldsymbol{U}$ on the left- and right-hand sides of (4). The latter is the easier to picture: the drift motion in the oscillating magnetic field $\boldsymbol{h}$ of the wave causes oscillations of the electrons which can be in phase with $\boldsymbol{e}$. The former effect is due to the terms $(\boldsymbol{U}\cdot\nabla)\boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{U}$ which may be pictured as an inertial force due to the combination of the drift and wave motions and which can also cause oscillations of the electrons in phase with $\boldsymbol{e}$. We now use equations (10) and (11) to discuss a special case in which these mechanisms occur.

If $\boldsymbol{U}$ is sufficiently small the wave may approximate to a circularly polarised wave travelling in the z-direction. Suppose that in a certain region we may put $e_- \approx 0$, $e_z \approx 0$, $n \approx 0$, $\partial e_+/\partial x \approx 0$ and $\partial e_+/\partial y \approx 0$, $u_- \approx 0$, $u_z \approx 0$ and the approximation (7) may be used in terms containing $\boldsymbol{U}$. The contribution from div $\boldsymbol{F}$ to the rate of exchange of energy then cancels the terms containing $\boldsymbol{U}$ on the left-hand side of (10), using $(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}^* \approx 0$ and div $(N\boldsymbol{U}) = 0$. Also using (1) and (7)

$$\boldsymbol{h}\wedge\boldsymbol{u}^* + \boldsymbol{h}^*\wedge\boldsymbol{u} \approx \frac{mc}{q}\frac{\omega+\Omega}{\omega}\nabla(\boldsymbol{u}_+\boldsymbol{u}_+^*)\,.$$

Then, since $\partial(u_+u_+^*)/\partial x \approx 0$ and $\partial(u_+u_+^*)/\partial y \approx 0$

$$\text{div}\,(\boldsymbol{P}+\boldsymbol{F}) \approx \frac{m}{4}\left\{\frac{\omega+\Omega}{\omega}NU_z\frac{\partial}{\partial z}(u_+u_+^*) - 2N\nu(u_+u_+^*)\right\} \qquad (12)$$

VOL. 4 1953

In this case the wave gains energy from the electron drift motion in any region where $U_z \dfrac{\partial}{\partial z}(u_+ u_+^*)$ is positive. Later in the more detailed work on a more restricted model, a better approximation will be obtained and the approximations justified, but the working is much more complicated, and it will be found that the important effect is described by (12). In the next section, in which the radiation from a sudden disturbance is discussed, an effect such as that described by (12) is shown to be capable of causing strong amplification. The following expression for the Poynting vector of an ordinary circularly polarised wave will be needed. Equations (5) and (8) yield

$$P_x = P_y = 0, \qquad P_z = \frac{ikc}{16\pi}(e_+ h_+^* - e_+^* h_+) \tag{13}$$

4. Radiation from a Transient Disturbance

The Laplace transform has been used by Jaeger and Westfold in a study of transients with applications to solar noise, and will now be adapted to discuss the problem of amplification. The Laplace transform $\bar{\xi}(p)$ of any variable $\xi(t)$ is defined by

$$\bar{\xi}(p) = \int_0^\infty e^{-pt}\xi \, dt \tag{14}$$

Here the Fourier transform is required; it is equal to $\bar{\xi}(i\omega)$ provided $\int_0^\infty \xi \, dt$ is absolutely convergent Carslaw and Jaeger (1949). We assume that this condition is satisfied when ξ is a component of the radiation field and define $\bar{n}(i\omega)$ by

$$\bar{n}(i\omega) = \int_0^\infty e^{-i\omega t}\,(N - N)\, dt. \tag{15}$$

$\bar{\boldsymbol{e}}$, $\bar{\boldsymbol{h}}$ and $\bar{\boldsymbol{u}}$ are similarly defined, the argument always being $i\omega$ in the remainder of this work; N, $\boldsymbol{E}$, $\boldsymbol{H}$ and $\boldsymbol{U}$ are the total electron density, electric field, magnetic field and electron velocity as in section 2. The equations for the Fourier transform are obtained from the fundamental field equations by multiplying by $e^{i\omega t}$ and integrating with respect to t from 0 to ∞; the integrals containing time derivatives are then transformed by partial integration; the resulting expression of the form $\left[e^{-i\omega t}\xi\right]_0^\infty$ reduces to $-\xi^{(0)}$ where $\xi^{(0)}$ is the value of ξ at $t = 0$. Thus from

$$\frac{\partial}{\partial t}(\boldsymbol{H} - \boldsymbol{H}) = -c \operatorname{curl}(\boldsymbol{E} - \boldsymbol{E}) \tag{16}$$

we obtain

$$i\omega\bar{\boldsymbol{h}} - (\boldsymbol{H} - \boldsymbol{H})^{(0)} = -c \operatorname{curl} \bar{\boldsymbol{e}} \tag{17}$$

In the other equations non-linear terms occur and will be approximated as in

$$\mathsf{N}\mathsf{U} - N\boldsymbol{U} \approx (\mathsf{N} - N)\boldsymbol{U} + N(\mathsf{U} - \boldsymbol{U})$$

the neglected term $(\mathsf{N} - N)(\mathsf{U} - \boldsymbol{U})$ being similar to the term $n\boldsymbol{u}$ neglected in (2). The equations corresponding to (2) and (4) are then

$$c \operatorname{curl} \bar{\boldsymbol{h}} - i\omega\bar{\boldsymbol{e}} + (\mathsf{E} - \boldsymbol{E})^{(0)} = 4\pi q(n\boldsymbol{U} + N\bar{\boldsymbol{u}}) \tag{18}$$

and

$$(i\omega + \nu)\bar{\boldsymbol{u}} - (\mathsf{U} - \boldsymbol{U})^{(0)} + (\boldsymbol{U} \cdot \nabla)\bar{\boldsymbol{u}} + (\bar{\boldsymbol{u}} \cdot \nabla)\boldsymbol{U}$$
$$= \frac{q}{m}(\bar{\boldsymbol{e}} + \bar{\boldsymbol{u}} \wedge \boldsymbol{H}/c + \bar{\boldsymbol{u}} \wedge \bar{\boldsymbol{h}}/c) \tag{19}$$

Thus $\bar{\boldsymbol{e}}$, $\bar{\boldsymbol{h}}$, $\bar{n}$ and $\bar{\boldsymbol{u}}$ satisfy equations (1) . . . (4) modified only by the addition of the terms representing the initial disturbance, that is the initial departure of E, H, N and U from the steady values $\boldsymbol{E}$, $\boldsymbol{H}$, N and $\boldsymbol{U}$. This investigation will now be restricted to a disturbance in E only, so that only equation (18) contains a term representing the initial disturbance. Further the disturbance is supposed to be strongly localised and will be represented by a Dirac δ-function, so that equations (1) . . . (4) are modified only at one value of z. Consider the components $\bar{e}_+$, $\bar{h}_+$ and $\bar{u}_+$ corresponding to an ordinary circularly polarised wave travelling in the z-direction. If

$$(\mathsf{E}_+ - E_+)^{(0)} = T\delta(z - s)$$

(18) gives

$$ic\frac{\partial \bar{h}_+}{\partial z} - i\omega\bar{e}_+ + T\delta(z - s) = 4\pi q N\bar{u}_+ \tag{20}$$

Equations (17) and (19) show that $\bar{e}_+$ and $\bar{u}_+$ are continuous at $z = s$, and hence from (20), $\bar{h}_+$ increases discontinuously at $z = s$ by the amount iT/c.

If now the boundary condition (A) is applied at $z = +\infty$, this will determine the ratio $\bar{h}_+/\bar{e}_+$ for any z greater than s, because (17), (18) and (19) are homogeneous except at $z = s$; it is convenient to write

$$\bar{h}_+/\bar{e}_+ = \mu$$

μ is related to the reflection coefficient looking in the direction of positive z and μ^{-1} is known as the wave impedance; also for a pure travelling wave μ is imaginary and $|\mu|$ is equal to the refractive index. Let μ_1 denote the limit of μ as z tends to s from above. Similarly the boundary condition (A) applied at $z = -\infty$ determines μ for $z < s$ and μ_2 the limit as z tends to s from below. Now at s, $\bar{e}_+$ is continuous and $\bar{h}_+$ increases discontinuously by the amount iT/c, hence

$$(\mu_1 - \mu_2)\bar{e}_+(s) = iT/c$$

and

$$|\bar{e}_+(s)|^2 = |T|^2/c^2\,|\mu_1 - \mu_2|^2 \tag{21}$$

(21) determines the intensity of the wave for a given disturbance described by T and shows that, if $|\mu_1 - \mu_2|$ were very small, the intensity would be very large,

VOL. 4 1953

that is resonance would occur. The value of $\mu_1 - \mu_2$ is therefore important; the real and imaginary parts of $\mu_1 - \mu_2$ require separate discussions.

From (13) we obtain

$$i\,l(\mu) = \frac{\mu^* - \mu}{2} = -\frac{8\pi i k}{c}\frac{P_z}{|\bar{e}_+|^2}$$

so that the imaginary part of μ is related to the Poynting vector. It should be noted that the kinetic energy flow F_z is continuous at $z = s$, because u_+ is continuous. If there were no source of energy other than the disturbance described by T, P_z would have to be positive for $z > s$ and negative for $z < s$; this would put a strong limitation on the possibility of resonance and the condition of exact resonance would require total loss-free reflection on both sides of the disturbance. In our problem however the electron velocity provides another source of energy and it is in principle possible for P_z to be negative in the region where z is a little greater than s in spite of the boundary condition (A); alternatively P_z could be positive in the region where z is a little less than s. This possibility is investigated in the remaining sections of this paper.

From (17)

$$i\omega\bar{h}_+ = -\,c\,\partial\bar{e}_+/\partial z$$

and hence

$$2R(\mu) = \mu + \mu^* = -\frac{c}{\omega}\frac{\partial}{\partial z}\log(\bar{e}_+\bar{e}_+{}^*)$$

Now in a region of real refractive index $\bar{e}_+\bar{e}_+{}^*$ oscillates and the condition that the real part of $(\mu_1 - \mu_2)$ vanishes is the same sort of condition as that which determines the frequency of any one-dimensional resonator. Since in our system the wavelength is much smaller than the dimensions of the system, this condition is very sensitive to small changes in various quantities, but the phenomenon of resonance will be significant, if it still occurs after averaging over the real part of $(\mu_1 - \mu_2)$ in an appropriate way: since $R(\mu)$ varies sinusoidally with z we put $R(\mu_1 - \mu_2) = m\cos\theta$ and average over θ; m is of the order of magnitude of the refractive index.

Denoting the imaginary part of $(\mu_1 - \mu_2)$ by l (21) gives

$$\langle|\bar{e}_+(s)|^2\rangle_{\mathrm{Av}} = \frac{|T|^2}{2\pi c^2}\int_0^{2\pi}\frac{d\theta}{l^2 + m^2\cos^2\theta}$$

which, on using the identity

$$\int_0^{2\pi}\frac{d\theta}{1 + \lambda^2\cos^2\theta} = 2\pi(1+\lambda^2)^{-\frac{1}{2}} \tag{22}$$

becomes

$$\langle|\bar{e}_+(s)|^2\rangle_{\mathrm{Av}} = \frac{|T|^2}{c^2 l\,(m^2 + l^2)^{\frac{1}{2}}} \tag{23}$$

Thus a sort of resonance can still occur, because $|\bar{e}_+(s)|^2$ would become very large if l became very small.

This method of averaging is not justifiable, if the disturbance occurs in an overdense region. Then $\bar{e}_+\bar{e}_+^*$ does not oscillate with z and there is a tendency for $\partial(\bar{e}_+\bar{e}_+^*)/\partial z$ to be negative for $z > s$ and positive for $z < s$.

The conclusion obtained from this section is that a large intensity can be obtained from a disturbance in a region of real refractive index, if the two values of the Poynting vector immediately on either side of s determined by the boundary condition (A) applied at $z = +\infty$ and $-\infty$ respectively are nearly equal. This can occur only if the wave absorbs energy from the medium, so that the effect is one of amplification as well as resonance.

5. THE EQUATION $(c^2D^2 + \alpha + i\beta + i\gamma D)e = 0$.

It will be necessary later to know the variation of the Poynting vector for a wave equation of the form

$$(c^2D^2 + \alpha + i\beta + i\gamma D)e = 0 \tag{24}$$

where D denotes $\partial/\partial z$, e is a transverse component of the electric field, and α, β and γ are functions of z. Let

$$\rho = e\,e^* \tag{25}$$

$$g = i(e^*De - eDe^*) \tag{26}$$

From (8) and (1) it is seen that P_z is $c^2g/8\pi\omega$, and from (24)

$$c^2Dg = (2\beta + \gamma D)\rho \tag{27}$$

It is convenient to introduce the new variables

$$e' = e\exp\left[\frac{i}{2c^2}\int\gamma\mathrm{d}z\right] \tag{28}$$

$$g' = g - \gamma\rho/c^2 \tag{29}$$

$$\alpha' = \alpha + \gamma^2/4c^2 \tag{30}$$

$$\beta' = \beta - \tfrac{1}{2}D\gamma \tag{31}$$

Then $e'e'^* = \rho$ and

$$(c^2D^2 + \alpha' + i\beta')e' = 0 \tag{32}$$

and

$$c^2Dg' = 2\beta'\rho \tag{33}$$

In considering the resonance condition of section 4, α, β and γ are given continuous functions of z. Two solutions of (24) are required: (1) for $z > s$, satisfying the boundary condition (A) at $z = +\infty$ and (2) for $z < s$, satisfying (A) at $z = -\infty$. The values of e at $z = s$ for the two solutions are to be made equal and then, if the two values of P_z at $z = s$ are nearly equal, resonance will occur. Thus the condition requires that the two values of g should be nearly equal, and, since ρ is continuous at $z = s$, that the two values of g' should be nearly equal.

It is convenient to write (33) in the integrated form

$$c^2g' = 2\int\beta'\rho\,\mathrm{d}z + K \tag{34}$$

VOL. 4 1953

The lower limit of the integral will be taken as $+\infty$ for (1) and $-\infty$ for (2) and the physical significance of the constant K can then be seen. For (1) the wave at large z is assumed to be a pure outgoing wave travelling in free space, so that $\alpha = \omega^2$, $\beta = \gamma = 0$ and

$$De = -i\omega e/c$$

whence

$$g = 2\omega\rho/c.$$

Since $\gamma = 0$, $g' = g$ and

$$K = 2c\omega\rho_0$$

where ρ_0 is the value of ρ far out in free space. For (2) the wave is supposed to pass through the sun and obviously decays very rapidly so that the constants corresponding to ρ_0 and K are both zero.

A somewhat more complicated situation occurs in the next section, but the discussion is similar. Any constant multiple of $NU\rho$ is continuous at $z = s$ and vanishes in free space; hence it can be added to g to give g' and the resonance condition still requires the two values of g' to be nearly equal. It then follows that the integral in (34) should contribute an amount nearly equal to $2\omega\rho_0/c$ to the difference between the values of g' at $z = s$. More precisely $\int_{-\infty}^{\infty} \beta'\rho \, dz$ in which ρ is given by the solution (1) for $z > s$ and by (2) for $z < s$, must be nearly equal to $c\omega\rho_0$.

6. Model with Axial Symmetry

In order to investigate the process of amplification in more detail it is necessary to consider a more definite model. A one-dimensional model would be too restrictive, because NU would then have to be a constant in order to satisfy the continuity equation. A convenient model is the axially symmetric one, which we now consider.

The fundamental equations (1) . . . (4) are developed by Mr. R. E. Loughhead (1953) in the following paper for a model specified as follows: $\boldsymbol{H}$ is uniform and directed in the z-direction and N depends on z only; in cylindrical coordinates r, ϕ and z, U_ϕ is zero and U_r and U_z are independent of ϕ and are odd and even functions of r respectively. The dependence on ϕ of the components of the wave is also restricted: they are to be proportional to $e^{-ik\phi}$.

All quantities depending on r are expanded in powers of r, the coefficients being indicated by suffices; for example

$$e_r = \sum_{n=0}^{\infty} e_{rn} r^n$$

and the zero order components of the wave correspond to an ordinary circularly polarised wave. Loughhead obtains the equations

$$4\pi i\omega N q u_{r0} = (\omega^2 + c^2\partial^2/\partial z^2) e_{r0} \tag{35}$$

$$\left(i(\omega + \Omega) + \nu - (2N)^{-1} \frac{\partial}{\partial z}(NU_{z0}) + U_{z0} \frac{\partial}{\partial z}\right) u_{r0}$$

$$= \frac{q}{m}\left(1 - i\frac{U_{z0}}{\omega}\frac{\partial}{\partial z}\right) e_{r0} + \frac{iqU_{z0}}{m\omega} e_{z1} \tag{36}$$

He then imposes the condition that the radial component $P_r + F_r$ of the total energy flow shall not contribute to the value of div $(\boldsymbol{P} + \boldsymbol{F})$ on the axis $r = 0$, that is $P_{r1} + F_{r1} = 0$. This means that any variation of $P_{z0} + F_{z0}$ with z must be due to energy exchange between the wave and the gas.

LOUGHHEAD shows that if

$$e_{z1} = (\xi + i\eta)e_{r0} \tag{37}$$

where ξ and η are real, this condition determines η, and that, under the conditions required in the remainder of this paper, a good enough approximation is

$$c^2 \frac{\partial}{\partial z}(\eta e_{r0} e_{r0}^*) = -4\pi\omega m\, u_{r0}\, u_{r0}^* \frac{\partial}{\partial z}(NU_{z0}) \tag{38}$$

One further equation is needed to determine the solution. It is consistent with LOUGHHEAD'S equations to specify ξ arbitrarily and later ξ will be taken to vanish for simplicity.

Now, if u_{r0} were eliminated from (35) and (36) they would yield a differential equation of third order; this would have two solutions representing radio waves and a third solution of quite a different sort with a wave velocity of the order of U_z. Under the conditions which are of interest, however, an approximate second order equation can be obtained for the radio waves. We suppose

$$|\boldsymbol{U}| \ll c \tag{39}$$

$$|N^{-1}\frac{\partial}{\partial z}(NU_{z0})| \ll \omega \tag{40}$$

$$\nu \ll \omega \tag{41}$$

Simple consideration also shows that, if the refractive index is less than 1, which is the case for ordinary polarisation, $u_{r0}^{-1}\partial u_{r0}/\partial z$ is not greater than ω/c. Hence the term $U_{z0}\partial u_{r0}/\partial z$ in (36) is not greater than $U_{z0}\omega u_{r0}/c$, which is much less than ωu_{r0} by (39). Thus neglecting the small terms in (36)

$$u_{r0} \approx -iqe_{r0}/m(\omega + \Omega) \tag{42}$$

and if this is now substituted in the small terms in (36) the resulting equation is

$$i(\omega + \Omega)u_{r0} = \frac{q}{m}\left\{\left[1 + \frac{i}{\omega + \Omega}\left(\nu - \frac{1}{2N}\frac{\partial}{\partial z}(NU_{z0}) - \frac{\Omega}{\omega}U_{z0}\frac{\partial}{\partial z}\right)\right]e_{r0} + \frac{iU_{z0}}{\omega}e_{z1}\right\} \tag{43}$$

Then, eliminating u_{r0} from (35) and (43) and using (37)

$$\left(c^2\frac{\partial^2}{\partial z^2} + \omega^2\right)e_{r0} = \frac{4\pi N\alpha^2\omega}{m(\omega + \Omega)}\left\{1 - \frac{U_{z0}(\eta - i\xi)}{\omega} + \frac{i}{\omega + \Omega}\left(\nu - \frac{1}{2N}\frac{\partial}{\partial z}(NU_{z0}) - \frac{\Omega}{\omega}U_{z0}\frac{\partial}{\partial z}\right)\right\}e_{r0}$$

VOL. 4 1953

This is of the form (24) with

$$\alpha = \omega^2\left[1 - \frac{N}{N_c}(1 - U_{z0}\eta/\omega)\right] \tag{44}$$

where

$$N_c = m\omega(\omega + \Omega)/4\pi q^2 \tag{45}$$

$$\beta = -\frac{\omega^2 N}{N_c}\left[\frac{U_{z0}\xi}{\omega} + \frac{\nu - (2N)^{-1}\partial(NU_{z0})/\partial z}{\omega + \Omega}\right] \tag{46}$$

$$\gamma = \omega\Omega N U_{z0}/N_c(\omega + \Omega) \tag{47}$$

It is seen that ξ and η appear only in the expressions for β and α respectively, and in (27) β appears but not α. Thus ξ contributes to Dg; but ξ can be specified arbitrarily, and for want of a better hypothesis we put $\xi = 0$. It should be noted that all the components of the wave except e_{r0}, $e_{\phi 0}$, vanish for a plane wave (i.e. when $\boldsymbol{U} = 0$) and hence, even if it were not taken to be zero, ξ could be regarded as being of at least first order in $\boldsymbol{U}$; then $U_{z0}\xi/\omega$ would be of second order in $\boldsymbol{U}$ and would not be likely to be important in (27).

The work of Section 5 cannot be taken straight over, because, as shown by Loughhead

$$P_{z0} = \frac{c^2}{8\pi\omega}(g + 2\eta\rho) \tag{48}$$

where now

$$\rho = e_{r0}e_{r0}^*$$

and

$$g = i(e_{r0}^* De_{r0} - e_{r0}De_{r0}^*)$$

We may put, in (38),

$$\frac{u_{r0}u_{r0}^*}{e_{r0}e_{r0}^*} \approx \left(\frac{q}{m(\omega + \Omega)}\right)^2 = \frac{\omega}{4\pi N_c m(\omega + \Omega)} \tag{49}$$

and then obtain from (48), (27) and (38) with (49)

$$\frac{\partial P_{z0}}{\partial z} = \frac{\partial}{\partial z}\left(\frac{\Omega N U_{z0}\rho}{8\pi N_c(\omega + \Omega)}\right) - \frac{\rho}{8\pi N_c(\omega + \Omega)}\left(2N\omega\nu + (\omega + \Omega)\frac{\partial(NU_{z0})}{\partial z}\right) \tag{50}$$

The earlier result (12) can be obtained using

$$F_{z0} = \tfrac{1}{2}NmU_{z0}u_{r0}u_{r0}^*$$

and using the approximation (49).

The resonance condition can now be expressed in the same form as in Section 5: $\int_{-\infty}^{\infty}\beta'\rho\,dz$ should be nearly equal to $c\omega\rho_0$, if β' is given by

$$\beta' = -\frac{N\omega^2\nu}{N_c(\omega + \Omega)} - \omega(2N_c)^{-1}\frac{\partial(NU_{z0})}{\partial z} \tag{51}$$

7. General Discussion

Since little is known of the state of the solar atmosphere over a sunspot, only a rough discussion will here be attempted of the situation in which the resonance

condition can be fulfilled. The condition as stated at the end of the last section requires that $\int_{-\infty}^{\infty} \beta' \rho \, dz$ should be nearly equal to $c\omega\rho_0$. This condition is obviously not impossible in principle and depends largely on β' which is given by (51).

Since the resonance condition can be satisfied only if the wave absorbs energy from the gas, and the effect of collisions, represented by the term involving ν, is a loss of energy to the gas, a necessary condition for resonance is that somewhere

$$\frac{\omega - \Omega}{2\omega N} D(NU) > \nu \tag{52}$$

where U is now written for U_{z0}.

Now ν is given approximately (SMERD, 1950) by the formula $40NT^{-3/2}$ secs^{-1} where T is the absolute temperature. Also from (39) $N_c = 3{\cdot}10^{-10}\,\omega(\omega + \Omega)$ if ω and Ω are measured in radians per second; and $\Omega = 1{\cdot}7{\cdot}10^{7}H$. Taking typical figures for the solar atmosphere and observed radio frequencies we find $\omega(\omega + \Omega) \sim 10^{18}$ and $T \sim 10^{6}$, which give $\nu \sim 10N/N_c$ secs^{-1}. We are not concerned with regions where N is much greater than N_c and hence (41) is easily satisfied; also (52) is not incompatible with (40). If the temperature of the emitting region in the solar atmosphere were higher than normal, ν would of course be smaller since it is proportional to $T^{-3/2}$.

Thus it appears that (52) can easily be satisfied, but it should be noted that ν is equal to the mean thermal velocity divided by the mean free path. The sort of conditions required by (52) are likely to have $DU > \nu$ and then the gas must depart appreciably from thermal conditions, that is the velocity distribution function for the gas will not be thermal. This does not invalidate our discussion, but shows that amplification will occur only in rather exceptional regions.

Leaving aside the loss due to collisions we wish to estimate the order of magnitude of $2\int_{-\infty}^{\infty} \beta' \rho dz$; from (51) the part involving U is

$$-\frac{\omega}{N_c}\int_{-\infty}^{\infty} \rho D(NU) dz \tag{53}$$

We discuss the contribution to this integral from a region where NU varies rapidly with z; (52) in fact shows that NU must change by a large amount in a distance small compared with the scale of the corona. If in this region NU changes from 0 to N_1U_1 where $N_1 \sim N_c$ and $U_1 \ll c$, the contribution to (53) from this region can be written $\omega U_1\rho_1$, and ρ will have the value ρ_1 at some point in this region. Then, if the amplification from such a region is to be sufficient to cause resonance, $\omega U_1\rho_1$ must be at least of order $2c\omega\rho_0$; that is

$$\frac{\rho_1}{\rho_0} > \frac{c}{U_1} \gg 1 \tag{54}$$

The important conclusion is that amplification must occur in a region where ρ is much greater than ρ_0. Then, if amplification occurs in a region where ρ reaches a value of ρ_1, U must become as large as $c\rho_0/\rho_1$. It is therefore important to know how large ρ becomes. In overdense regions, where $N > N_c$, ρ can increase by a very large factor, but, if overdense regions are involved, the discussion is more

VOL. 4 1953

complicated and it is difficult to estimate the relative importance of various effects; consequently we shall consider only a region in which $N < N_c$, although we shall suppose that N passes through the value N_c at the boundary of this region. The relevant equation can be taken as (32) and, since the question under investigation is whether ρ becomes large enough for the effect of β' to become important, β' can be put equal to zero in this investigation, for, if the effect of β' is important, the question is answered. (32) then becomes

$$(c^2D^2 + \alpha')e' = 0$$

and the W.K.B. approximation for this equation has been given by BREMMER (1951). In this approximation

$$\rho = \rho_0 \omega \alpha'^{-1/2} \tag{55}$$

and the approximation is valid so long as α' varies by only a small factor in one local wavelength. Since the local wavelength is $2\pi c\alpha'^{1/2}$, the condition of validity can be written

$$\left| 2\pi c\alpha'^{-3/2} D\alpha' \right| \ll 1 \tag{56}$$

The approximation (55) will give the right order of magnitude until the left hand side of (56) is of the order of 1, so that, if α' passes through zero, ρ will reach the order of magnitude ρ_m given by

$$m_0 \omega (2\pi c D\alpha')^{-1/3} \tag{57}$$

Now α' is given by (30) and we suppose that α changes more rapidly than $\gamma^2/4c^2$. We also suppose that DN is roughly constant and equal to kN_c in the region near the point where $\alpha' = 0$, so that k^{-1} is the scale length for N. Then from (44), $\left| D\alpha' \right| \sim k\omega^2$ and

$$\rho_m \sim \rho_0 \frac{\omega}{2\pi ck} \tag{58}$$

Now $2\pi c/\omega$ is the free space wavelength, say 1 metre; k^{-1} may be of the order of the scale height in the corona, which is about $5 \cdot 10^4$ km. Consequently ρ_m may be several hundred times larger than ρ_0, and from (54) amplification can be obtained with velocities of the order of 1000 km/sec.

The argument given above does not prove that electron drift velocities as large as 1000 km/sec are necessary for amplification. For instance if α' has a minimum for some value of z, there will be a small range of frequencies for which the minimum value of α' is very small. For these frequencies ρ/ρ_0 will become very large in the region near the minimum of α' and amplification may be possible with electron velocities substantially smaller than 1000 km/sec. No further detailed discussion will be given at present. The aim of this paper has been to demonstrate the possibility in principle of amplification resulting from electronic motion and the question whether it is actually responsible for any solar storms still remains open.

7. ACKNOWLEDGEMENTS

This work has been carried out during the tenure of a Senior Research Fellowship established with the help of Commonwealth Research Grants to the University of Sydney.

The author is indebted to Professor V. A. BAILEY for his encouragement and helpful advice and to Dr. K. C. WESTFOLD and Mr. R. E. LOUGHHEAD for useful discussions.

REFERENCES

BAILEY, V. A.	1948	Aust. J. Sci. Res. **1**, 351
	1950	Phys. Rev. **78**, 428
BREMMER, H.	1951	Communication on Pure and Applied Mathematics, IV, 105
BUNEMAN, O.	1951	Proceedings of the Conference on Dynamics of Ionised Media at University College, London, April
CARSLAW, H. S. and JAEGER, J. C.	1949	Operational Methods in Applied Mathematics, Chap. XII (Second edition, Oxford)
JAEGER, J. C. and WESTFOLD, K. C.	1949	Aust. J. Sci. Res. A, **2**, 322
LOUGHHEAD, R. E.	1953	J. Atmosph. Terrs. Phys. **4**, 163
SMERD, S. F.	1950	P.I.E.E. **97**, Part III, 447
TWISS, R. Q.	1950	Phys. Rev. **80**, 767
	1951	Nature **169**, 185

VOL.
4
1953

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 163 to 174. Pergamon Press Ltd., London

The equations for a problem arising in Dungey's investigation of Bailey's theory of sunspot noise.

R. E. Loughhead

School of Physics, The University of Sydney, N.S.W.

(*Received* 9 *July* 1953)

ABSTRACT

In the preceding investigation (p. 148) of wave amplification, Dr. J. W. Dungey has paid particular attention to the possibility of wave amplification in an axially symmetric model of the solar atmosphere over a sunspot. In the present paper the differential equations of a small radiation field in a non-uniform ionised gas are shown to possess particular solutions, which, along the axis of symmetry, correspond to ordinary circularly polarised waves. By taking one of these solutions subject to the condition that there be no net radial flow of wave energy either into, or away from, the axis, the electric vector $\boldsymbol{e}_0$ of the wave along the axis and the associated electron velocity $\boldsymbol{u}_0$ are found to be related by the equations

$$4\pi i\omega Nq\boldsymbol{u}_0 = \left\{\omega^2 + c^2\frac{\partial^2}{\partial z^2}\right\}\boldsymbol{e}_0$$

$$\left\{i(\omega+\Omega) + \nu - \frac{1}{2N}\frac{\partial}{\partial z}(NU_{z0}) + U_{z0}\frac{\partial}{\partial z}\right\}\boldsymbol{u}_0 = \frac{q}{m}\left\{1 - \frac{U_{z0}}{\omega}\left(\eta + i\frac{\partial}{\partial z}\right)\right\}\boldsymbol{e}_0,$$

where η obeys the equation

$$c^2\frac{\partial}{\partial z}(\eta\boldsymbol{e}_0\cdot\boldsymbol{e}_0^*) = -4\pi m\omega\frac{\partial}{\partial z}(NU_{z0})\boldsymbol{u}_0\cdot\boldsymbol{u}_0^*$$

These equations form the starting point of Dungey's investigation of wave amplification in the axially symmetric model.

1. Introduction

The problem discussed in this paper arises from the preceding investigation p. 148 by Dr. J. W. Dungey of wave amplification in the non-uniform solar atmosphere over a sunspot. It is there shown that a large intensity can be obtained from a disturbance in a region of real refractive index, provided that a certain resonance condition is satisfied. To examine the process in detail it is necessary to consider a particular model of the solar atmosphere over a sunspot.

As it is impossible to construct a one-dimensional model in which the equation of continuity of the electrons in the ionised atmosphere is satisfied everywhere, we must consider a model which is at least two-dimensional. A convenient model, upon which the investigation of this paper is based, is the particular axially symmetric model described in Section 3. In the following sections the equations of a small radiation field in an ionised gas are applied to this model, and the differential equations governing the radiation field along the axis of symmetry of the model are derived by means of the method of Taylor expansion.

2. THE FUNDAMENTAL EQUATIONS

The differential equations governing a small radiation field in a non-uniform ionised gas have been developed by Dr. J. W. DUNGEY from the fundamental equations of BAILEY'S electro-magneto-ionic theory, subject to certain simplifying assumptions. For convenience we shall use the notations adopted in Section 2 of the preceding paper, where the symbols $\boldsymbol{e}$, $\boldsymbol{h}$, $\boldsymbol{u}$ and n denote respectively the electric vector, the magnetic vector, the associated electron velocity and the associated electron density of the radiation field. In addition the symbols $\boldsymbol{H}$, $\boldsymbol{U}$ and N denote respectively the steady magnetic field, the steady mean electron velocity and the steady electron density of the ionised gas; q is the electronic charge, $-4{\cdot}8 \times 10^{-10}$ e.s.u., m is the mass of an electron and c is the velocity of light in free space. With this notation and with $\boldsymbol{e}$, $\boldsymbol{h}$, $\boldsymbol{u}$ and n taken proportional to $e^{i\omega t}$, the differential equations obtained by DUNGEY may be expressed as follows:

$$i\omega \boldsymbol{h} = -c \operatorname{curl} \boldsymbol{e} \tag{1}$$

$$c \operatorname{curl} \boldsymbol{h} - i\omega \boldsymbol{e} = 4\pi q(N\boldsymbol{u} + n\boldsymbol{U}) \tag{2}$$

$$i\omega n = -\operatorname{div}(N\boldsymbol{u} + n\boldsymbol{U}) \tag{3}$$

$$(i\omega + \nu)\boldsymbol{u} - \boldsymbol{\Omega} \times \boldsymbol{u} + (\boldsymbol{u} \cdot \operatorname{grad})\boldsymbol{U} + (\boldsymbol{U} \cdot \operatorname{grad})\boldsymbol{u} = \frac{q}{m}\left\{\boldsymbol{e} + \frac{1}{c}\boldsymbol{U} \times \boldsymbol{h}\right\} \tag{4}$$

where

$$\boldsymbol{\Omega} = -\frac{q}{mc}\boldsymbol{H} \tag{5}$$

In place of (3) we may use the relation

$$\operatorname{div} \boldsymbol{e} = 4\pi q n, \tag{6}$$

which follows from (2) and (3).

By eliminating $\boldsymbol{h}$ between (1) and (4) we obtain the equation

$$(i\omega + \nu)\boldsymbol{u} - \boldsymbol{\Omega} \times \boldsymbol{u} + (\boldsymbol{u} \cdot \operatorname{grad})\boldsymbol{U} + (\boldsymbol{U} \cdot \operatorname{grad})\boldsymbol{u} = \frac{q}{m}\boldsymbol{e} + i\frac{q}{m\omega}\boldsymbol{U} \times \operatorname{curl} \boldsymbol{e} \tag{7}$$

For the purpose of expressing (7) by a set of three equations in cylindrical co-ordinates it is convenient to write it in the form

$$(i\omega + \nu)\boldsymbol{u} - \boldsymbol{\Omega} \times \boldsymbol{u} + \operatorname{grad}(\boldsymbol{u} \cdot \boldsymbol{U}) - \boldsymbol{U} \times \operatorname{curl} \boldsymbol{u} - \boldsymbol{u} \times \operatorname{curl} \boldsymbol{U}$$
$$= \frac{q}{m}\boldsymbol{e} + i\frac{q}{m\omega}\boldsymbol{U} \times \operatorname{curl} \boldsymbol{e} \tag{8}$$

Similarly, if we eliminate $\boldsymbol{h}$ and n between equations (1), (2) and (6) we obtain the result

$$4\pi i\omega N q \boldsymbol{u} = \omega^2 \boldsymbol{e} - i\omega(\operatorname{div} \boldsymbol{e})\boldsymbol{U} - c^2 \operatorname{curl} \operatorname{curl} \boldsymbol{e} \tag{9}$$

The investigation which follows is based on equations (8) and (9).

3. APPLICATION OF THE FUNDAMENTAL EQUATIONS TO AN AXIALLY SYMMETRIC MODEL OF THE SOLAR ATMOSPHERE

The fundamental vector equations (8) and (9) will now be applied to a two-dimensional model, which is axially symmetric about the z-axis. This model is specified as follows:

1. The steady magnetic field $\boldsymbol{H}$ is uniform and parallel to the z-axis.

VOL. 4 1953

2. The steady electron density N and the electron collision frequency ν depend on z alone.

3. If U_r, U_ϕ and U_z denote the components of the steady mean electron velocity, specified with respect to cylindrical co-ordinates r, ϕ and z, then U_ϕ is zero, U_r and U_z are independent of ϕ and are respectively odd and even functions of r.

If we write $\Omega = |\boldsymbol{\Omega}|$ in conformity with DUNGEY's notation, then

$$\Omega_z = -k\Omega, \tag{10}$$

where $k = -1$ if $\Omega_z > 0$ and $k = 1$ if $\Omega_z < 0$. We note that $k^2 = 1$.

When the components of $\boldsymbol{e}$ and $\boldsymbol{u}$ are taken proportional to $e^{-ik\phi}$, solutions exist, which, along the z-axis, correspond to an ordinary circularly polarised wave. Then, by expressing (8) and (9) as a set of six equations in cylindrical co-ordinates, we obtain the following simultaneous partial differential equations:

$$\left\{(i\omega+\nu)+\frac{\partial U_r}{\partial r}+U_r\cdot\frac{\partial}{\partial r}+U_z\cdot\frac{\partial}{\partial z}\right\}u_r-k\Omega u_\phi+\frac{\partial U_r}{\partial z}u_z$$
$$=\frac{q}{m}\left\{1-i\frac{U_z}{\omega}\frac{\partial}{\partial z}\right\}e_r+i\frac{q}{m}\cdot\frac{U_z}{\omega}\frac{\partial e_z}{\partial r} \tag{11}$$

$$k\Omega u_r+\left\{(i\omega+\nu)+\frac{U_r}{r}+U_r\frac{\partial}{\partial r}+U_z\frac{\partial}{\partial z}\right\}u_\phi$$
$$=k\frac{q}{m}\cdot\frac{U_r}{\omega}\frac{e_r}{r}+\frac{q}{m}\left\{1-\frac{i}{\omega}\left[U_r\left(\frac{1}{r}+\frac{\partial}{\partial r}\right)+U_z\frac{\partial}{\partial z}\right]\right\}e_\phi+k\frac{q}{m}\cdot\frac{U_z}{\omega}\cdot\frac{e_z}{r} \tag{12}$$

$$\frac{\partial U_z}{\partial r}\cdot u_r+\left\{(i\omega+\nu)+\frac{\partial U_z}{\partial z}+U_z\cdot\frac{\partial}{\partial z}\right\}u_z$$
$$=i\frac{q}{m}\cdot\frac{U_r}{\omega}\frac{\partial e_r}{\partial z}+\frac{q}{m}\left\{1-i\frac{U_r}{\omega}\frac{\partial}{\partial r}\right\}e_z \tag{13}$$

$$4\pi i\omega Nqu_r=\left\{\omega^2-i\omega U_r\left(\frac{1}{r}+\frac{\partial}{\partial r}\right)-c^2\left(\frac{1}{r^2}-\frac{\partial^2}{\partial z^2}\right)\right\}e_r$$
$$-k\left\{\frac{\omega U_r}{r}-ic^2\left(\frac{1}{r^2}+\frac{1}{r}\frac{\partial}{\partial r}\right)\right\}e_\phi-\left\{i\omega U_r\frac{\partial}{\partial z}+c^2\frac{\partial^2}{\partial r\partial z}\right\}e_z \tag{14}$$

$$4\pi i\omega Nqu_\phi=ikc^2\frac{\partial}{\partial r}\left(\frac{e_r}{r}\right)+\left\{\omega^2-c^2\left(\frac{1}{r^2}-\frac{1}{r}\frac{\partial}{\partial r}-\frac{\partial^2}{\partial r^2}-\frac{\partial^2}{\partial z^2}\right)\right\}e_\phi+ikc^2\frac{1}{r}\frac{\partial e_z}{\partial z} \tag{15}$$

$$4\pi i\omega Nqu_z=-\left\{i\omega U_z\left(\frac{1}{r}+\frac{\partial}{\partial r}\right)+c^2\left(\frac{1}{r}\frac{\partial}{\partial z}+\frac{\partial^2}{\partial r\partial z}\right)\right\}e_r-\frac{k}{r}\left\{\omega U_z-ic^2\frac{\partial}{\partial z}\right\}e_\phi$$
$$+\left\{\omega^2+c^2\left(\frac{1}{r}\frac{\partial}{\partial r}+\frac{\partial^2}{\partial r^2}\right)-\frac{c^2}{r^2}-i\omega U_z\cdot\frac{\partial}{\partial z}\right\}e_z \tag{16}$$

To investigate the form of the radiation field governed by these equations we shall express the components of $\boldsymbol{e}$ and $\boldsymbol{u}$ as Taylor series in r, writing

$$\left.\begin{aligned} e_r &= \sum_{n=0}^{\infty} e_{rn} \cdot r^n, \text{ etc.} \\ u_r &= \sum_{n=0}^{\infty} u_{rn} \cdot r^n, \text{ etc.} \end{aligned}\right\} \tag{17}$$

where e_{rn}, $e_{\phi n}$, e_{zn}, u_{rn}, $u_{\phi n}$ and u_{zn} are functions of ϕ, z and t. Moreover, as we do not know the functional dependence of U_r and U_z on r, we shall express these quantities as Taylor series in r; thus

$$U_r = \sum_{n=1}^{\infty} U_{rn} r^n; \qquad U_z = \sum_{n=0}^{\infty} U_{zn} r^n, \tag{18}$$

where, because U_r and U_z are respectively odd and even functions of r, we have

$$\left.\begin{aligned} U_{r,2l} &= 0 \\ U_{z,2l+1} &= 0 \end{aligned}\right\} \tag{19}$$

for all positive integral values of l. If we substitute (17) and (18) into the set of equations (11) . . . (16), and equate the coefficients of the various powers of r, we obtain a set of differential recurrence relations. The recurrence relations of nth order may be stated as follows:

$$\begin{aligned} (i\omega + \nu)u_{rn} + \sum_{l=0}^{n}\left\{\frac{\partial u_{r,n-l}}{\partial z} U_{zl}\right\} + \sum_{l=1}^{n+1}\left\{(n+1)u_{r,n+1-l}U_{rl}\right\} - k\Omega u_{\phi n} \\ + \sum_{l=1}^{n}\left\{u_{z,n-l}\frac{\partial U_{rl}}{\partial z}\right\} = \frac{q}{m} e_{rn} - i\frac{q}{m\omega}\sum_{l=0}^{n}\left\{\frac{\partial e_{r,n-l}}{\partial z} U_{zl}\right\} \\ + i\frac{q}{m\omega}\sum_{l=0}^{n+1}\left\{(n+1-l)e_{z,n+1-l}U_{zl}\right\} \end{aligned} \tag{20}$$

$$\begin{aligned} k\Omega u_{rn} + (i\omega + \nu)u_{\phi n} + \sum_{l=0}^{n}\left\{\frac{\partial u_{\phi,n-l}}{\partial z} U_{zl}\right\} + \sum_{l=1}^{n+1}\left\{(n+2-l)u_{\phi,n+1-l}U_{rl}\right\} \\ = k\frac{q}{m\omega}\sum_{l=1}^{n+1}\left\{e_{r,n+1-l}U_{rl}\right\} + \frac{q}{m}\cdot e_{\phi n} - i\frac{q}{m\omega}\sum_{l=0}^{n}\left\{\frac{\partial e_{\phi,n-l}}{\partial z} U_{zl}\right\} \\ - i\frac{q}{m\omega}\sum_{l=1}^{n+1}\left\{(n+2-l)e_{\phi,n+1-l}U_{rl}\right\} + k\cdot\frac{q}{m\omega}\sum_{l=0}^{n+1}\left\{e_{z,n+1-l}U_{zl}\right\} \end{aligned} \tag{21}$$

$$\begin{aligned} \sum_{l=0}^{n-1}\left\{l\cdot u_{r,n+1-l}U_{zl}\right\} + (i\omega + \nu)u_{zn} + \sum_{l=0}^{n}\left\{u_{z,n-l}\frac{\partial U_{zl}}{\partial z} + \frac{\partial u_{z,n-l}}{\partial z} U_{zl}\right\} \\ + \sum_{l=1}^{n+1}\left\{(n+1-l)u_{z,n+1-l}U_{rl}\right\} \\ = i\frac{q}{m\omega}\sum_{l=1}^{n}\left\{\frac{\partial e_{r,n-l}}{\partial z} U_{rl}\right\} + \frac{q}{m} e_{zn} - i\frac{q}{m\omega}\sum_{l=1}^{n+1}\left\{(n+1-l)e_{z,n+1-l}U_{rl}\right\} \end{aligned} \tag{22}$$

VOL. 4 1953

$$4\pi i\omega N q u_{rn} = \left\{\omega^2 + c^2\frac{\partial^2}{\partial z^2}\right\}e_{rn} - i\omega\sum_{l=1}^{n+1}\left\{(n+2-l)e_{r,n+1-l}U_{rl}\right\}$$
$$- k\omega\sum_{l=1}^{n+1}\left\{e_{\phi,n+1-l}U_{rl}\right\} - i\omega\sum_{l=1}^{n}\left\{\frac{\partial e_{z,n-l}}{\partial z}U_{rl}\right\}$$
$$- c^2\left\{e_{r,n+2} - ik(n+3)e_{\phi,n+2}\right\} - c^2(n+1)\frac{\partial e_{z,n+1}}{\partial z} \qquad (23')$$

$$4\pi i\omega N q u_{\phi n} = \left\{\omega^2 + c^2\frac{\partial^2}{\partial z^2}\right\}e_{\phi n} + ikc^2(n+1)\{e_{r,n+2} - ik(n+3)e_{\phi,n+2}\}$$
$$+ ikc^2\frac{\partial e_{z,n+1}}{\partial z} \qquad (24')$$

$$4\pi i\omega N q u_{zn} = -i\omega\sum_{l=0}^{n+1}\{(n+2-l)e_{r,n+1-l}U_{zl}\}$$
$$- k\omega\sum_{l=0}^{n+1}\{e_{\phi,n+1-l}U_{zl}\} + \omega^2 e_{zn} - i\omega\sum_{l=0}^{n}\left\{\frac{\partial e_{z,n-l}}{\partial z}U_{zl}\right\}$$
$$- c^2(n+2)\frac{\partial e_{r,n+1}}{\partial z} + ikc^2\frac{\partial e_{\phi,n+1}}{\partial z} + c^2(n+3)(n+1)e_{z,n+2} \qquad (25)$$

As the six recurrence relations corresponding to any assigned n do not involve the variables $e_{r,n+2}$ and $e_{\phi,n+2}$ independently, but only through the linear combination

$$\alpha_{n+2} = e_{r,n+2} - ik(n+3)e_{\phi,n+2} \qquad (26)$$

it is convenient to replace (23′) and (24′) by the equivalent relations

$$4\pi i\omega N q\{(n+1)u_{rn} - iku_{\phi n}\} = \left\{\omega^2 + c^2\frac{\partial^2}{\partial z^2}\right\}\{(n+1)e_{rn} - ike_{\phi n}\}$$
$$- i\omega(n+1)\sum_{l=1}^{n+1}\{(n+2-l)e_{r,n+1-l}U_{rl}\} - k\omega(n+1)\sum_{l=1}^{n+1}\{e_{\phi\,n+1-l}U_{rl}\}$$
$$- i\omega(n+1)\sum_{l=1}^{n}\left\{\frac{\partial e_{z,n-l}}{\partial z}U_{rl}\right\} - c^2 n(n+2)\frac{\partial e_{z,n+1}}{\partial z} \qquad (23)$$

$$4\pi i\omega N q u_{\phi n} = \left\{\omega^2 + c^2\frac{\partial^2}{\partial z^2}\right\}e_{\phi n} + ikc^2(n+1)\alpha_{n+2} + ikc^2\frac{\partial e_{z,n+1}}{\partial z} \qquad (24)$$

Equation (23) is obtained by eliminating α_{n+2} between equations (23′) and (24′), and equation (24) is simply equation (24′) rewritten with the help of (26). We shall henceforth refer to equations (20), (21), (22), (23), (24) and (25) as the general nth order recurrence formulae.

Because the coefficients in the expansions of U_r and U_z satisfy the relations (19) it is easily seen by inspection that, for every positive integer n, no recurrence relation contains coefficients from both of the following sets:

1. Even coefficients in the expansions of e_r, e_ϕ, u_r and u_ϕ, and odd coefficients in the expansions of e_z and u_z.

2. Odd coefficients in the expansions of e_r, e_ϕ, u_r and u_ϕ, and even coefficients in the expansions of e_z and u_z.

Consequently, solutions of the homogeneous linear equations (11) . . . (16) may be constructed by forming linear combinations of solutions that correspond to taking either all the coefficients in set (1), or all the coefficients in set (2), equal to zero. For the purpose of this paper it is sufficient to consider solutions in which

$$\left.\begin{aligned} e_{r,2l+1} &= 0 \\ e_{\phi,2l+1} &= 0 \\ e_{z,2l} &= 0 \end{aligned}\right\} \quad \text{and} \quad \left.\begin{aligned} u_{r,2l+1} &= 0 \\ u_{\phi,2l+1} &= 0 \\ u_{z,2l} &= 0 \end{aligned}\right\} \tag{27}$$

for all positive integral values of l. It therefore follows that the recurrence relations (20), (21), (23) and (24) are satisfied identically for all odd positive integers n, and that the recurrence relations (22) and (25) are satisfied identically for all even positive integers n. Because these equations are satisfied identically, they need not be considered further.

VOL. 4 1953

4. Form of the Radiation Field

If we take $n = -2$ in (24) and $n = 0$ in (20), (21), (23) and (24) we obtain the following equations:

$$e_{r0} = ike_{\phi 0} \tag{28}$$

$$4\pi i\omega Nq(u_{r0} - iku_{\phi 0}) = \left\{\omega^2 + c^2\frac{\partial^2}{\partial z^2}\right\}(e_{r0} - ike_{\phi 0}) - i\omega U_{r1}(e_{r0} - ike_{\phi 0}) \tag{29}$$

$$4\pi i\omega Nqu_{\phi 0} = \left\{\omega^2 + c^2\frac{\partial^2}{\partial z^2}\right\}e_{\phi 0} + ikc^2\alpha_2 + ikc^2\frac{\partial e_{z1}}{\partial z} \tag{30}$$

$$\left\{(i\omega + \nu) + U_{r1} + U_{z0}\frac{\partial}{\partial z}\right\}u_{r0} - k\Omega u_{\phi 0} = \frac{q}{m}\left\{1 - i\frac{U_{z0}}{\omega}\frac{\partial}{\partial z}\right\}e_{r0} + i\frac{q}{m}\cdot\frac{U_{z0}}{\omega}\cdot e_{z1} \tag{31}$$

$$k\Omega u_{r0} + \left\{(i\omega + \nu) + U_{r1} + U_{z0}\frac{\partial}{\partial z}\right\}u_{\phi 0} = \frac{q}{m}\left\{1 - i\frac{U_{r1}}{\omega} - i\frac{U_{z0}}{\omega}\frac{\partial}{\partial z}\right\}e_{\phi 0}$$

$$+ k\cdot\frac{q}{m}\cdot\frac{U_{r1}}{\omega}e_{r0} + k\cdot\frac{q}{m}\cdot\frac{U_{z0}}{\omega}e_{z1} \tag{32}$$

Equations (28) and (29) yield the relation

$$u_{r0} = iku_{\phi 0} \tag{33}$$

Consequently, equations (31) and (32) are not independent, as each can be transformed into the other with the aid of (28) and (33), and therefore equations (28) . . . (32) reduce to the two independent equations

$$4\pi i\omega N q u_{r0} = \left\{\omega^2 + c^2 \frac{\partial^2}{\partial z^2}\right\} e_{r0} - c^2 \alpha_2 - c^2 \frac{\partial e_{z1}}{\partial z} \tag{34}$$

$$\left\{i(\omega + \Omega) + \nu + U_{r1} + U_{z0} \frac{\partial}{\partial z}\right\} u_{r0} = \frac{q}{m}\left\{1 - i \frac{U_{z0}}{\omega} \frac{\partial}{\partial z}\right\} e_{r0} + i \frac{q}{m} \cdot \frac{U_{z0}}{\omega} \cdot e_{z1} \tag{35}$$

involving the four independent variables e_{r0}, u_{r0}, α_2 and e_{z1}.

The six additional equations obtained by putting $n = 1$ in (22) and (25) and $n = 2$ in (20), (21), (23) and (24) involve six additional unknowns u_{z1}, e_{r2} or $e_{\phi 2}$, u_{r2}, $u_{\phi 2}$, e_{z3} and α_4, as α_2 is a linear combination of e_{r2} and $e_{\phi 2}$. Moreover, the six further equations obtained by putting $n = 3$ in (22) and (25) and $n = 4$ in (20), (21), (23) and (24) involve six additional unknowns u_{z3}, e_{r4} or $e_{\phi 4}$, u_{r4}, $u_{\phi 4}$, e_{z5} and α_6, as α_4 is a linear combination of e_{r4} and $e_{\phi 4}$. Similarly, all successive sets of six additional equations taken in accordance with the above procedure will involve in every case six additional unknowns. Two further equations are therefore required, and these will be obtained from physical considerations.

In imposing the first condition we are guided by the fact that terms of the order of U^2 are neglected in DUNGEY's investigation, which means, on the whole, that the results obtained may be regarded as correct to the first approximation for small $\boldsymbol{U}$. When $\boldsymbol{U} = 0$ the recurrence equations are still indeterminate, and if the equations (34) and (35), which relate the components of $\boldsymbol{e}$ and $\boldsymbol{u}$ along the z-axis, are to correctly describe the propagation of plane circularly polarised waves along the z-axis in the limit as $\boldsymbol{U} \to 0$, we must take

$$\alpha_2 = -\frac{\partial e_{z1}}{\partial z} \tag{36}$$

This is evident from an inspection of (34) and (35) as the term $i \cdot \frac{q}{m} \cdot \frac{U_{z0}}{\omega} \cdot e_{z1}$ in (35) vanishes in the limit as $\boldsymbol{U} \to 0$, but the terms $-c^2\alpha_2$ and $-c^2 \frac{\partial e_{z1}}{\partial z}$ in (34) do not.

Since $e_{z0} = 0$ and $u_{z0} = 0$ by (27), equations (28) and (33) show that the radiation along the z-axis corresponds to an ordinary circularly polarised wave. The component e_{r0} of the electric vector of this wave is related to the component u_{r0} of the associated electron velocity through equations (34) and (35). In view of (36) equation (34) reduces to the form

$$4\pi i\omega N q u_{r0} = \left\{\omega^2 + c^2 \frac{\partial^2}{\partial z^2}\right\} e_{r0} \tag{37}$$

Equation (35) may be expressed in a more convenient form if we note

$$U_{r1} = -\frac{1}{2N}\frac{\partial}{\partial z}(NU_{z0}), \tag{38}$$

which is easily verified by using the equation of continuity, div $N\boldsymbol{U} = 0$, for the electrons in the gas and by remembering that $U_{r0} = 0$. Further it is convenient for the later discussion to express e_{z1} in terms of e_{r0} by writing

$$e_{z1} = (\xi + i\eta)e_{r0} \tag{39}$$

where ξ, η are real quantities yet to be determined. Then equation (35) may be written as

$$\left\{i(\omega + \Omega) + \nu - \frac{1}{2N}\frac{\partial}{\partial z}(NU_{z0}) + U_{z0}\frac{\partial}{\partial z}\right\}u_{r0} = \frac{q}{m}\left\{1 - i\frac{U_{z0}}{\omega}\frac{\partial}{\partial z}\right\}e_{r0}$$

$$-\frac{q}{m}\cdot\frac{U_{z0}}{\omega}(\eta - i\xi). \tag{40}$$

The field along the z-axis is not, however, determined by equations (37) and (40) alone, because equation (40) contains the two unknown quantities ξ and η. Although we could introduce the single arbitrary (complex) relation still at our disposal by assigning ξ and η the value zero for simplicity, the solution so obtained would not be physically significant in DUNGEY's discussion. The consideration that must guide the introduction of the second condition to be imposed on our solution is the vital question of the mean rate of flow of wave energy into the z-axis. This question will be discussed in the following section of this paper.

VOL. 4 1953

5. FLOW OF WAVE ENERGY INTO THE AXIS OF SYMMETRY

In the present paper it is essential to eliminate the possibility that wave amplification along the z-axis might be due merely to a radial flow of wave energy into the axis, and not due to the mechanism suggested by BAILEY, in which the radiation field would absorb energy from the drift motion of the electrons. To counter this possible objection we impose the condition that there be no net radial flow of wave energy either into, or away from, the z-axis.

The total mean flow of wave energy is

$$\bar{\boldsymbol{P}} + \bar{\boldsymbol{F}}, \tag{41}$$

where $\bar{\boldsymbol{P}}$ is the time-averaged value of the Poynting vector of the radiation field, and $\bar{\boldsymbol{F}}$ is the time-averaged flow of the kinetic energy of the electronic wave motion induced by the radiation field. If there is to be no net radial flow of wave energy into or away from the z-axis, then the contribution of the radial component of the total mean energy flow (41) to the value of div $(\bar{\boldsymbol{P}} + \bar{\boldsymbol{F}})$ on the z-axis must vanish. That is, we require

$$\underset{r\to 0}{Lt}\left\{\frac{1}{r}\frac{\partial}{\partial r}(r\bar{P}_r)\right\} + \underset{r\to 0}{Lt}\left\{\frac{1}{r}\frac{\partial}{\partial r}(r\bar{F}_r)\right\} = 0 \tag{42}$$

DUNGEY shows that $\bar{\boldsymbol{P}}$ is given by the equation

$$\bar{\boldsymbol{P}} = \frac{c}{16\pi}(\boldsymbol{e}^* \times \boldsymbol{h} + \boldsymbol{e} \times \boldsymbol{h}^*) \tag{43}$$

where, by equation (1),

$$\boldsymbol{h} = \frac{ic}{\omega} \operatorname{curl} \boldsymbol{e} \tag{44}$$

By expressing the radial component $\bar{P}_r$ of $\bar{\boldsymbol{P}}$ and the components of $\boldsymbol{h}$ as Taylor series in r, it follows from (43) that, in a self-explanatory notation,

$$\bar{P}_{r0} = \frac{c}{16\pi}\{e_{\phi 0}{}^* h_{z0} - e_{z0}{}^* h_{\phi 0} + e_{\phi 0} h_{z0}{}^* - e_{z0} h_{\phi 0}{}^*\},$$

and

$$\bar{P}_{r1} = \frac{c}{16\pi}\{e_{\phi 0}{}^* h_{z1} + e_{\phi 1}{}^* h_{z0} - e_{z0}{}^* h_{\phi 1} - e_{z1}{}^* h_{\phi 0} + e_{\phi 0} h_{z1}{}^* + e_{\phi 1} h_{z0}{}^* - e_{z0} h_{\phi 1}{}^* - e_{z1} h_{\phi 0}{}^*\}$$

Because e_{z0} and $e_{\phi 1}$ are zero and $e_{r0} = ike_{\phi 0}$, these expressions reduce to the forms

$$\bar{P}_{r0} = \frac{ikc}{16\pi}\{e_{r0}{}^* h_{z0} + e_{r0} h_{z0}{}^*\}$$

$$\bar{P}_{r1} = \frac{ikc}{16\pi}\{e_{r0}{}^* h_{z1} - e_{r0} h_{z1}{}^*\} - \frac{c}{16\pi}\{e_{z1}{}^* h_{\phi 0} + e_{z1} h_{\phi 0}{}^*\}$$

By equation (44),

$$h_\phi = \frac{ic}{\omega}\left\{\frac{\partial e_r}{\partial z} - \frac{\partial e_z}{\partial r}\right\}$$

and

$$h_z = \frac{ic}{\omega}\left\{\frac{\partial e_\phi}{\partial r} + \frac{e_\phi}{r} + ik\frac{e_r}{r}\right\}$$

Therefore,

$$h_{\phi 0} = \frac{ic}{\omega}\left\{\frac{\partial e_{r0}}{\partial z} - e_{z1}\right\}$$

$$h_{z0} = 0$$

and

$$h_{z1} = -k \cdot \frac{c}{\omega} \cdot \alpha_2,$$

where

$$\alpha_2 = e_{r2} - ik \cdot 3e_{\phi 2}$$

We thus find

$$\bar{P}_{r0} = 0 \tag{45}$$

and

$$\bar{P}_{r1} = \frac{ic^2}{16\pi\omega}\left\{(e_{r0}\alpha_2{}^* - e_{r0}{}^*\alpha_2) - \left(\frac{\partial e_{r0}}{\partial z} \cdot e_{z1}{}^* - \frac{\partial e_{r0}{}^*}{\partial z} \cdot e_{z1}\right)\right\} \tag{46}$$

On account of (36) equation (46) becomes

$$\bar{P}_{r1} = \frac{ic^2}{16\pi\omega}\frac{\partial}{\partial z}(e_{r0}{}^{*}e_{z1} - e_{r0}e_{z1}{}^{*})$$

and, if we introduce the relation (39), namely

$$e_{z1} = (\xi + i\eta)e_{r0},$$

reduces to the form

$$\bar{P}_{r1} = -\frac{c^2}{8\pi\omega}\frac{\partial}{\partial z}(\eta e_{r0}e_{r0}{}^{*}) \tag{47}$$

Now, since $\bar{P}_{r0} = 0$, it is easily seen that the first limit in (42) has the value $2\bar{P}_{r1}$, and hence, by equation (47), this limit is found to be

$$\underset{r\to 0}{Lt}\left\{\frac{1}{r}\frac{\partial}{\partial r}(r\bar{P}_r)\right\} = -\frac{c^2}{4\pi\omega}\frac{\partial}{\partial z}(\eta e_{r0}e_{r0}{}^{*}) \tag{48}$$

It is noteworthy that this result involves η but not ξ.

To evaluate the second limit in (42) we commence with the expression (cf. DUNGEY, Sec. 2)

$$\bar{\boldsymbol{F}} = \tfrac{1}{4}mN(\boldsymbol{u}\cdot\boldsymbol{u}^{*})\boldsymbol{U} + \tfrac{1}{4}m\{\boldsymbol{U}\cdot(n^{*}\boldsymbol{u} + n\boldsymbol{u}^{*})\}\boldsymbol{U} + \tfrac{1}{4}mN(\boldsymbol{U}\cdot\boldsymbol{u}^{*})\boldsymbol{u} + \tfrac{1}{4}mN(\boldsymbol{U}\cdot\boldsymbol{u})\boldsymbol{u}^{*} + \tfrac{1}{8}mU^{2}(n^{*}\boldsymbol{u} + n\boldsymbol{u}^{*}). \tag{49}$$

for the mean flow of the kinetic energy of the wave motion of the electrons, terms of the fourth order in the small quantities $\boldsymbol{u}$ and n being neglected. Since our fundamental equations (8) and (9) are correct only to the first order of these small quantities, it would be pointless to retain these fourth order terms in the expression (49). By equation (6), we have

$$n = (4\pi q)^{-1}\,\mathrm{div}\,\boldsymbol{e} \tag{50}$$

If n is expanded as a power series in r, the coefficients n_0 and n_1 in this expansion are easily found from (50) to be

$$\left.\begin{aligned} n_0 &= 0 \\ n_1 &= (4\pi q)^{-1}\left\{3e_{r2} - ike_{\phi 2} + \frac{\partial e_{z1}}{\partial z}\right\} \end{aligned}\right\} \tag{51}$$

Then, using (51) and remembering that $U_{r0} = 0$, it can be shown from (49) by repeated application of DE L'HÔPITAL'S limit theorem that the second limit in (42) has the value

$$\underset{r\to 0}{Lt}\left\{\frac{1}{r}\frac{\partial}{\partial r}(r\bar{F}_r)\right\} = 2mNU_{r1}u_{r0}u_{r0}{}^{*} + \tfrac{1}{2}mNU_{z0}(u_{r0}u_{z1}{}^{*} + u_{r0}{}^{*}u_{z1}) + \frac{mU_{z0}{}^{2}}{16\pi q}\left[u_{r0}\left\{3e_{r2}{}^{*} + ike_{\phi 2}{}^{*} + \frac{\partial e_{z1}{}^{*}}{\partial z}\right\} + u_{r0}{}^{*}\left\{3e_{r2} - ike_{\phi 2} + \frac{\partial e_{z1}}{\partial z}\right\}\right] \tag{52}$$

VOL. 4 1953

By considering the recurrence relation

$$\left\{i\omega + \nu + U_{r1} + \frac{\partial U_{z0}}{\partial z} + U_{z0}\frac{\partial}{\partial z}\right\}u_{z1}$$

$$= i\frac{q}{m}\cdot\frac{U_{r1}}{\omega}\cdot\frac{\partial e_{r0}}{\partial z} - 2\cdot U_{z2}u_{r0} + \frac{q}{m}\left\{1 - i\frac{U_{r1}}{\omega}\right\}e_{z1} \qquad (53)$$

obtained by taking $n = 1$ in (22), it may be shown that, when $\xi = 0$, the second term on the right hand side of (52) is several orders of magnitude smaller than the first term under the conditions in which the results of this paper are applied in DUNGEY'S discussion. The third term on the right hand side of (52) is rather complicated and cannot be estimated accurately. However, as it is of the third order in U_{z0}, it is probably small by comparison with the first term, which is of the first order in U_{z0}, and will be neglected in the present discussion. Then, combining equations (48) and (52), we find that the second condition to be imposed on our solution is expressed approximately by the equation

$$c^2\frac{\partial}{\partial z}(\eta e_{r0}e_{r0}^*) = -4\pi m\omega\frac{\partial}{\partial z}(NU_{z0})u_{r0}u_{r0}^* \qquad (54)$$

provided that we also put

$$\xi = 0. \qquad (55)$$

We are at liberty to choose ξ and η to satisfy (54) and (55) as we have yet to impose the second relation needed to obtain a determinate set of recurrence equations. Such a relation among complex quantities is equivalent to two separate relations among real quantities, and hence both ξ and η are at our disposal.

In the solution corresponding to this choice of ξ and η the component e_{r0} of the electric vector of the circularly polarised wave along the z-axis and the component u_{r0} of the associated electron velocity are determined by the equations

$$4\pi i\omega Nqu_{r0} = \left\{\omega^2 + c^2\frac{\partial^2}{\partial z^2}\right\}e_{r0} \qquad (56)$$

$$\left\{i(\omega + \Omega) + \nu - \frac{1}{2N}\frac{\partial}{\partial z}(NU_{z0}) + U_{z0}\frac{\partial}{\partial z}\right\}u_{r0} = \frac{q}{m}\left\{1 - \frac{U_{z0}}{\omega}\left(\eta + i\frac{\partial}{\partial z}\right)\right\}e_{r0} \qquad (57)$$

where η obeys the approximate relation

$$c^2\frac{\partial}{\partial z}(\eta e_{r0}e_{r0}^*) = -4\pi m\omega\frac{\partial}{\partial z}(NU_{z0})u_{r0}u_{r0}^* \qquad (54)$$

These equations form the starting point of DUNGEY'S investigation of wave amplification in the axially symmetric model.

ACKNOWLEDGEMENTS

This work has been carried out during the tenure of a Research Studentship established with the help of a University Research Grant from the University of Sydney.

The author wishes to express his gratitude to Professor V. A. BAILEY and to Dr. J. W. DUNGEY for guidance and assistance.

APPENDIX

In DUNGEY's discussion an expression is required for the mean flow of radiation energy along the z-axis. With the aid of equations (43) and (44) we readily obtain

$$\bar{P}_{z0} = \frac{ic^2}{8\pi\omega}\left\{\left(e_{r0}^* \frac{\partial e_{r0}}{\partial z} - e_{r0} \frac{\partial e_{r0}^*}{\partial z}\right) + (e_{r0}e_{z1}^* - e_{r0}^* e_{z1})\right\}$$

and hence, by equation (39), we have

$$\bar{P}_{z0} = \frac{c^2}{8\pi\omega}\left\{i\left(e_{r0}^* \frac{\partial e_{r0}}{\partial z} - e_{r0} \frac{\partial e_{r0}^*}{\partial z}\right) + 2\eta e_{r0}e_{r0}^*\right\} \tag{58}$$

as the required expression for the mean flow of radiation energy along the z-axis.

VOL. 4 1953

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 175 to 183. Pergamon Press Ltd., London

Solar tides in the F_2 region from the study of night-time critical frequencies

A. A. Weiss

Department of Physics, University of Adelaide

(*Received* 27 *May* 1953)

ABSTRACT

Night-time critical frequency variations at 25 ionospheric stations are analysed by season and by latitude for semi-diurnal solar tidal terms. The amplitude and phase of the vertical drift velocity of electrons so found are consistent with accepted tidal theory, and an estimate of the height-gradient of the vertical drift is obtained. Two parallel analyses are made, on the alternative assumptions that decay proceeds according to a recombination law or to an attachment law. The low values found for the decay coefficients preclude any decision as to which of these two decay processes is actually operative.

1. Introduction

It is now well established that the structure of the F-region is influenced by solar semi-diurnal components of tidal origin (Martyn, 1947, 1948; and Mitra, 1950, 1951). However, vertical tidal displacement of the electrons of an ionized region whose rate of ion-production follows the simple Chapman law is quite inadequate to resolve the complexities exhibited by the daytime F-region. Two further significant processes which have been considered are a possible diurnal cycle in the temperature of the F-region, and a decay coefficient which decreases with increasing height. With the aid of these additional processes a reasonable synthesis of the daytime F-region may be effected.

Night-time conditions in the F-region are very much simpler. The F_1-region has disappeared; it is unlikely that any appreciable temperature changes persist especially in the latter hours of darkness; the height of the F_2-region does not vary widely and the decay coefficient is therefore essentially uniform. It appears then to be a justifiable simplification to regard the night-time F_2-region as an isothermal ionized region with a uniform decay coefficient, subject to semi-diurnal tidal displacements. Under these conditions an analytic approach to the tidal problem becomes possible.

This paper is an attempt at the delineation of the structure of the vertical tidal pattern at the level of the F_2-region, by latitude and by season, from analysis of night-time critical frequencies for a large number of ionospheric stations. The appearance of tidal components in the (instantaneous) electron density implies a vertical gradient in the amplitude and/or phase of the drift, and by comparison of the tidal terms obtained from the study of critical frequencies with the somewhat meagre data available from analyses of height variations in h'_{F_2}, some useful conclusions may be drawn regarding the vertical gradient of the drift.

Values of the decay coefficient also follow from the analysis.

2. Method of Analysis

Under the assumptions, already mentioned, of an isothermal region with a uniform recombination coefficient, the continuity equation for electrons, in the absence of ion-production, is

$$\frac{\partial N}{\partial t} = -\alpha N^2 - \frac{\partial}{\partial z}(Nw)$$

where w is the semi-diurnal vertical drift velocity, measured positive upwards, in units of scale height H, and z is the reduced height $(h - h_0)/H$. At the level of maximum electron density N_m, this becomes

$$\frac{\mathrm{d}N_m}{\mathrm{d}t} = -\alpha N_m^2 - N_m \frac{\partial w}{\partial z} \tag{1}$$

For w, with height-gradient specified by y and σ, we may write

$$w = w_0 e^{-yz} \sin(2\omega t + \sigma z + \phi) \tag{2}$$

w_0 and ϕ will be functions of both season and latitude. Moreover, as the variation of the level of N_m is small, it is sufficient to identify this level with the datum level $z = 0$, adjusting w_0 and ϕ as required. Then, at the level $z = 0$, (1) becomes

$$\frac{\mathrm{d}N_m}{\mathrm{d}t} = -\alpha N_m^2 + AN_m \sin(2\omega t + \theta)$$

where $A = w_0\sqrt{y^2 + \sigma^2}$

$$\theta = \phi + \lambda; \quad \tan\lambda = -\sigma/y$$

In terms of critical frequency f^0 and a time scale in hours,

$$\frac{2}{f^0}\frac{\mathrm{d}f^0}{\mathrm{d}t} = -1{\cdot}24 \times 10^4 \alpha f^{02} + A \sin(2\omega t + \theta) \tag{3}$$

Equation (3) refers to a region whose ionization decays according to a recombination law. As considerable attention has been directed to the possibility that for the F_2-region decay proceeds according to an attachment (linear) law, the observational data have been analyzed also assuming this law of decay, and the two treatments run parallel through the paper. The equation for the attachment law, corresponding to equation (3) for the recombination law, is

$$\frac{2}{f^0}\frac{\mathrm{d}f^0}{\mathrm{d}t} = -\beta + A \sin(2\omega t + \theta) \tag{4}$$

The integral of (4) has been given by Mitra (1951), but the integral of (3) cannot be obtained in a useful form.

These equations have therefore been solved in differential form by numerical evaluation of $\mathrm{d}f^0/\mathrm{d}t$ at intervals of 3 hours. For a given station three values of $\mathrm{d}f^0/\mathrm{d}t$ (taken at 21 hrs, 00 hrs ≡ midnight and 03 hrs) suffice to determine the tidal parameters A and θ and the decay coefficient $\alpha(\beta)$. Without additional information regarding the height-gradient of the total drift it is not possible to

VOL. 4 1953

evaluate w_0 and ϕ which specify the drift velocity at the datum level, although the following considerations show that the phase angle ϕ can be fixed within reasonable limits. If $y > 0$ and $\sigma = 0$ (no phase gradient) $\phi = \theta$; if $y = 0$ (no amplitude gradient), $\phi = \theta \pm \pi/2$ according as $\sigma \gtrless 0$; and in the general case with $y \neq 0$; $\sigma \neq 0$, $|\phi - \theta| < \pi/2$. Alternatively, knowledge of w_0 and ϕ from independent data such as height variations will permit the evaluation of y and σ and hence a complete specification of the vertical tidal drift.

3. Data used in the Analysis

The data analyzed are the monthly averages of critical frequencies during the 8 hours of darkness from 20 hrs. to 04 hrs., for 25 stations for the period 1947–48, near sunspot maximum. The data for each station have been further averaged by seasons, each of 4 months duration: summer (S), equinox (E) and winter (W).

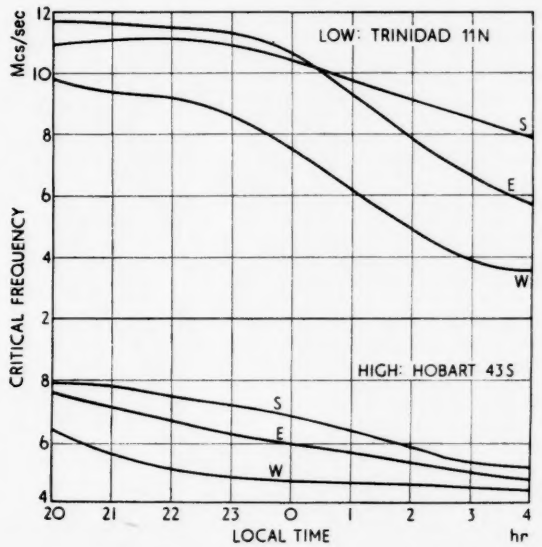


Fig. 1. Night-time critical frequency curves for typical low and high latitude stations.

Eighteen of the stations lie north of the geographic equator and seven to the south. A grouping by geographic latitude was found desirable, the categories and the number of stations falling in each being: Low, 0–20° (5); Medium, 21–34° (8); and High $\gg$ 35° (12).

Typical night-time variations of critical frequency with time at low and high latitude stations, for each of the three seasons, are shown in Fig. 1.

4. The Semi-diurnal Tidal Terms

Figs. 2 and 3 present, in harmonic dial form, the amplitude A and phase θ of the tidal term $A \sin(2\omega t + \theta)$ in (3) and (4), for different seasons and latitude groups. Fig. 2 applies if decay proceeds by attachment, and for the present attention will

be confined to this case, as giving less scatter and better trends than the case in which decay proceeds by recombination (Fig. 3).

With two exceptions, the scatter in the individual harmonic dials is large. Part of this scatter no doubt originates in the simplifying assumptions made. The phase angle θ is composite, depending on the parameters y and σ which specify the tidal gradient and upon the choice of a datum level. Therefore, whilst the phase of the eastwards component of the horizontal driving wind may remain

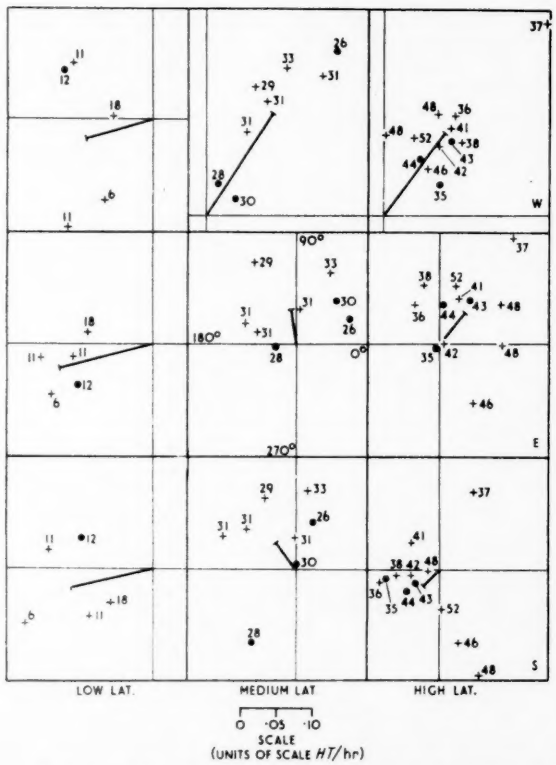


Fig. 2. Amplitude and phase of the tidal term $A \sin(2\omega t + \theta)$ for different seasons and latitude groups when decay follows an attachment law. Full lines are group means.
+ North latitudes · South latitudes.

independent of latitude (or at most reverse), the structure of the vertical tide, which is influenced by the height of the datum level, by geomagnetic asymmetry and by damping, may show considerable variation from station to station, consequently scatter in the derived values of A and θ is not surprising. It may be significant that the scatter is least at medium and high latitudes in winter, when the F_2-region shows closest approach to simple Chapman behaviour.

At the equinoxes, the phase angles θ for individual stations fall roughly into two overlapping groups, 150° apart, the overlap extending over the medium latitude range. Similar grouping is found, with much less scatter, in winter, with the cross-over point displaced towards the equator. No large seasonal change is

VOL. 4 1953

apparent at low latitudes. Seasonal vectors for the attachment case are drawn in Fig. 4 (full lines).

These annual (equinoctial) and seasonal components, whose summed amplitude is $w_0\sqrt{y^2+\sigma^2}$ and whose phase we may tentatively identify with the phase angle ϕ of the drift velocity at the datum level, are similar in form to the components proposed by MARTYN (1948) to account for critical frequency data for some Australian stations. There are differences, however, in the relative phases

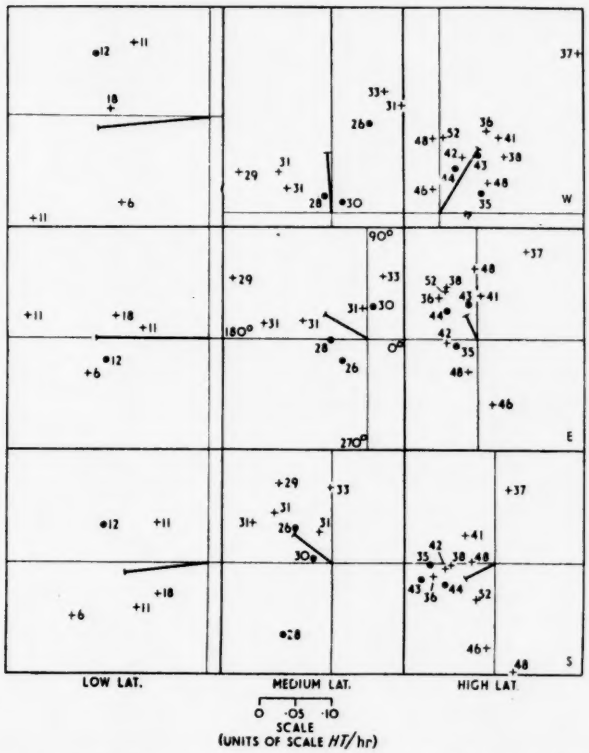


Fig. 3. Amplitude and phase of the tidal term A sin $(2\omega t + \theta)$ for different seasons and latitude groups when decay follows a recombination law. Full lines are group means.
+ North latitudes. South latitudes.

of the annual and seasonal terms—here either in or out of phase, rather than in phase quadrature as found by MARTYN; and in the smaller amplitude of the seasonal term in the present case.

The treatment so far has been confined to the tidal terms which must be present if decay proceeds by attachment. If however, decay proceeds by recombination, we find (Fig. 3) a larger scatter, an annual term in which the phase difference between low and high latitudes is reduced to about 90°, seasonal components in phase quadrature with annual components, and a different cross-over latitude in winter. Despite these differences there is a close correspondence between the behaviour of the tidal terms whichever law of decay is chosen. In

view of the low values found for the decay coefficient (section 5) this is not surprising, and in fact in summer when the decay coefficient is particularly small ($\alpha \sim 2 \times 10^{-11}$) the tidal terms become independent of the law of decay.

It is appropriate to insert at this point a remark upon seasonal tidal terms. The annual term may be denoted by $w_a \sin 2\omega t$, where w_a is a function of latitude and reverses sign at the node. Two alternative types of seasonal term are then available (a) $w_s(\delta) \sin (2\omega t + \xi)$ and (b) $w_s \sin (2\omega t + \Omega t + \zeta)$ where δ is the solar

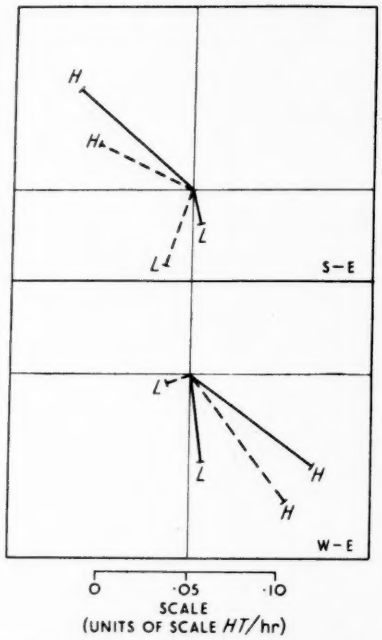


Fig. 4. Seasonal tidal terms for low and high latitude groups. Twice the scale of Figs. 2 and 3.
———— attachment — — — — — recombination.

declination and $\Omega t = 30°$/month. Chapman (1919) has shown that terms of type (a) with $\xi = 0$ or π will arise in the daily magnetic variations from a semi-diurnal wind system and a diurnal and seasonal variation of conductivity, and are characterized by a seasonal shift of the centres of the current sheets which produce the magnetic variations. As it is reasonable to suppose that the vertical electron drift reverses sign at the latitude of the centre of the current sheet, which shifts by some 10° between summer and winter, a seasonal term of the type (*a*) with $\xi = 0$ (or π) may well be present in the vertical tide. On the other hand $\xi \neq 0$ implies a continuous variation of the phase of the total tide with latitude, with no nodal point. Provided the seasonal term of type (*a*) is somewhat larger than the annual, the total tide is in phase opposition in summer and winter, and is identical in amplitude and phase at the two equinoxes. A seasonal term of type (*b*), proposed by Mitra (1951), gives a total tide whose phase advances by about 90° every three months and on this hypothesis one would expect the behaviour of

VOL. 4 1953

the F_2-region to be markedly different at the two equinoxes. That such behaviour was not observed in the data analyzed for high latitudes must be construed in favour of a seasonal tidal term of the type (*a*), viz., $w_s(\delta) \sin(2\omega t + \xi)$.

As indicated in Section 2, the values of A and θ, when considered in conjunction with independent data relating to height variations, will yield useful information regarding y and σ which specify the gradient of the vertical drift. WEISS (1953) has shown that a height gradient when w_0 is as large as $\frac{1}{2}$ has little influence on the motion close to the datum level, and there is ample evidence that the actual drift velocities are much less than this. It is therefore sufficient to consider the maximum upwards displacement of h^m as lagging 3 hours on the maximum upwards velocity. The further assumption $y \neq 0$, $\sigma = 0$ applied to the data of Figs. 2 and 3 leads to the time of maximum upwards displacements listed in Table 1. If $\sigma > 0$. $y = 0$, all the times of Table 1 are to be advanced by 3 hours. The

Table 1

Season / *Latitude*	*Attachment*		*Recombination*	
	S	W	S	W
Low	23·30	23·30	00·00	00·00
Medium	01·30	04·00	01·00	03·00
High	22·30	04·00	23·00	04·00

height data relating to the F_2-region collected by MITRA (1950) show a maximum upwards displacement shortly after midnight, and for Canberra earlier in summer than in winter. The times of Table 1 will fall into agreement with the height data provided $\sigma > 0$ and is not too small.

If the amplitudes of the $h^m_{F_2}$ variations quoted by MITRA are typical, w_0 is of the order 10 km/hr or less. Since $A \sim 0{\cdot}1$ scale height $= 7$ km, $\sqrt{y^2 + \sigma^2} = A/w_0 \sim 1$ and y and σ are of order unity. This is of course a very rough estimate, but one which is consistent with the absence of any large night-time variations in $h^m_{F_2}$.

5. THE DECAY COEFFICIENTS

The values of the decay coefficients found whilst solving equations (3) and (4) for the tidal parameters are plotted in Fig. 5.

The values of the recombination coefficient α for the three seasons, averaged over all latitudes, are: S $1{\cdot}6 \times 10^{-11}$, E $3{\cdot}1 \times 10^{-11}$ and W $6{\cdot}9 \times 10^{-11}$ cm^3 sec^{-1} respectively. Of the three abnormally large coefficients found, two relate to stations (White Sands and San Francisco) which do not show the expected seasonal change from summer to winter in the tide. Recombination coefficients for winter and equinox increase with latitude, linear regression coefficients found by the

method of least squares being $0{\cdot}12 \times 10^{-11}$ and $0{\cdot}08 \times 10^{-11}$/° latitude respectively. In summer the recombination coefficient is essentially independent of latitude. These values of α are much smaller than those usually accepted for the night-time F_2-region, as e.g. in the model atmosphere of Bates and Massey (1946) for which $\alpha = 3 \times 10^{-10}$ at night.

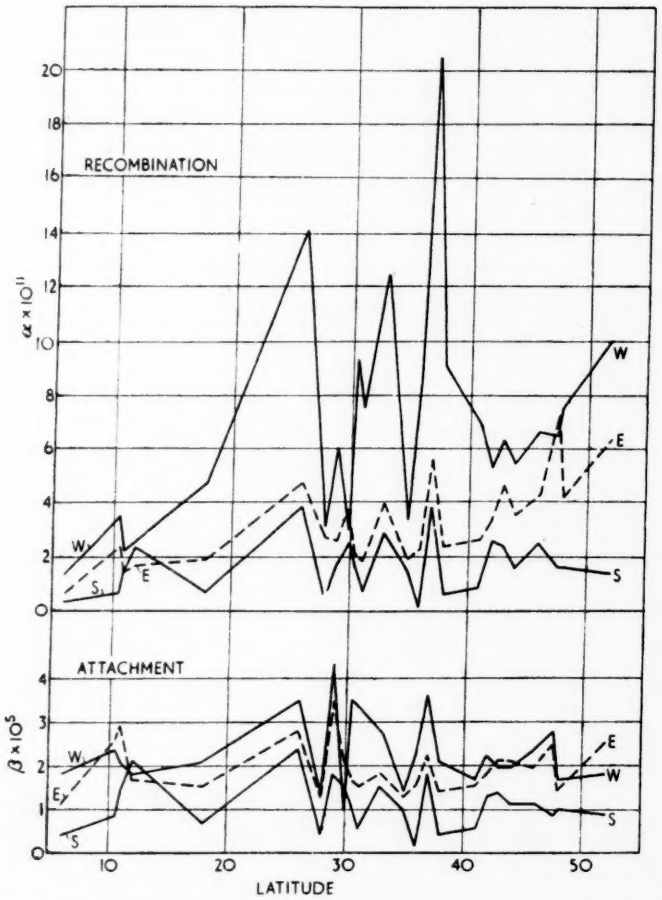


Fig. 5. Attachment and recombination coefficients for different seasons.

For decay by attachment, the attachment coefficient β is independent of latitude at all seasons. Values of β averaged over all latitudes are: S $1{\cdot}1 \times 10^{-5}$, E $1{\cdot}9 \times 10^{-5}$ and W $2{\cdot}4 \times 10^{-5}$ sec^{-1}.

It is interesting to observe that with values of A of the order of $\frac{1}{10}$ scale height/hr $= 1{\cdot}4 \times 10^{-5}$ scale height/sec, and values of $\beta \sim 2 \times 10^{-5}$/sec, the contributions to df^0/dt by tidal processes and by decay are of roughly equal importance.

VOL. 4 1953

6. Conclusion

It was hoped at the start of this investigation that the parallel treatment of decay by attachment and decay by recombination might furnish a firm basis for decision as to which of the two decay processes is actually operative in the F_2-region. That this expectation has not been realized is largely due to the small values found for the decay coefficients, which make it virtually impossible to distinguish between the two alternative decay processes over the short period of time (6 hours) examined. The smaller scatter in the tidal terms for different stations and a tidal pattern more in keeping with accepted tidal theory may be construed in favour of a linear law of decay but the evidence can scarcely be held to be conclusive.

The consideration of tidal terms by latitude groups is rather crude but suffices to establish the important influence of the gradient of the vertical drift upon the critical frequency. Further progress in evaluating the gradient of the drift rests upon simultaneous examination of height and critical frequency data for individual ionospheric stations, and such an investigation is planned.

Acknowledgements—The basic material for this survey was drawn from publications of the Australian Ionospheric Prediction Service. It is a pleasure to acknowledge the continued interest shown by Professor L. G. H. Huxley in this project. The author's thanks are also due to Mr. W. G. Elford for helpful discussion.

References

Bates, D. R., and Massey, H. S. W.	1946	Proc. Roy. Soc. A. **187,** 261
Chapman, S.	1919	Phil. Trans. London A. **218,** 1
Martyn, D. F.	1947	Proc. Roy. Soc. A. **189,** 241
	1948	*ibid* **194,** 445
Mitra, A. P.	1950	Ind. J. Phys. **24,** 387
	1951	J. Atmosph. Terr. Phys. **1,** 286
Weiss, A. A.	1953	J. Atmosph. Terr. Phys. **3,** 30

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 184 to 203. Pergamon Press Ltd., London

Semidiurnal currents and electron drifts in the ionosphere

J. A. FEJER

Telecommunication Research Laboratory, South African Council for Scientific and Industrial Research, Johannesburg, S.A.

(*Received* 13 *June* 1953)

ABSTRACT

The differential equations of the dynamo theory for the terrestrial ionosphere are solved numerically under simplifying assumptions.

The calculated current densities are used to estimate tidal velocities in the ionosphere necessary to produce the quiet day magnetic variations. A tidal amplification of about 60 compared with velocities on the ground is calculated.

The calculated velocities and phases of vertical electron drifts are in reasonable agreement with experimental determinations of lunar tidal variations in measured virtual height of the different ionospheric layers.

The calculated velocities and phases of horizontal electron drifts fit observations on long duration meteor echoes but the calculated phases are opposed to those obtained from fading measurements. The fading measurements appear therefore to refer to air movements and not to electron drifts.

1. INTRODUCTION

The chief cause of small regular fluctuations of the earth's magnetic field is believed to be an upper atmospheric current system. BALFOUR STEWART (1882) and SCHUSTER (1906) advanced the view that these currents are probably due to tidal movements of conducting air across the earth's magnetic field. CHAPMAN (1919) showed that the (height integrated) conductivity of the atmosphere was about $2{\cdot}5 \times 10^{-5}$ e.m.u. if the tidal speed was at all heights the same as it was on the ground. As knowledge of the conducting region of the atmosphere, the ionosphere, increased through the use of new experimental methods employing radio waves, it was hoped that the dynamo theory would be confirmed. APPLETON (1937) and others found, however, that the conductivity of the ionosphere calculated from radio measurements was far below that estimated by CHAPMAN (1919) as necessary to explain the quiet day magnetic variations.

TAYLOR (1936) and PEKERIS (1937) showed that the tidal oscillations of the atmosphere could occur in such a way as to cause the wind velocities at great heights to be very much greater than those on the ground. Recently this has been experimentally confirmed (SALZBERG and GREENSTONE, 1951; PHILLIPS, 1952). The velocities obtained, while about a hundred times higher than those on the ground, are certainly not as large as the theories of Taylor and Pekeris would allow them to be in the absence of damping. Recent studies of ionospheric conductivity (COWLING, 1945; BATES and MASSEY, 1951) have indicated that this tidal amplification of 100 is insufficient to explain the magnetic variations and that an amplification of at least 6000 is required.

This discrepancy led the author to re-examine the fundamental principles involved in the problem of ionospheric conductivity. When work on this paper was started, he was not aware that at least two others (HIRONO, 1952; MARTYN, 1952) had attacked the same problem. While the horizontal conductivities

VOL. 4 1953

derived by these authors agree with similar expressions in this paper, the method of solving the differential equation for the current density differs essentially from the methods of HIRONO and MARTYN and leads to different results.

2. CONDUCTIVITY OF THE IONOSPHERE

In the ionosphere electrons and ions execute their thermal motion in the presence of a magnetic field. If an electric field is present as well, there will be a mean drift of charged particles relative to the neutral particles superimposed on the thermal motions. The magnitude and direction of the drift velocity depend on the electric field, the magnetic field, the ratio of charge to mass of the particles and their collision frequency with the neutral air molecules.

Let the z' axis of a right handed cartesian coordinate system x', y', z' point in the direction of the magnetic field of flux density B. An electric field in the x', z' plane is assumed with components $E_{z'}$, $E_{x'}$ parallel and at right angles to the magnetic field. The existence of different kinds of charged particles together with a much larger number of neutral particles is assumed. The mean velocity of the neutral particles is assumed to be zero.

Let the charge of the nth kind of charged particle be e_n, its mass m_n, its collision frequency with neutral particles ν_n and its gyro-frequency $\omega_n = e_n B/m_n$. Then the mean velocity components $v_{x'n}$, $v_{y'n}$, $v_{z'n}$ of the nth charged particle relative to the neutral particles will be given by

$$v_{x'n} = E_{x'}\tau_{1n} \tag{1a}$$

$$v_{y'n} = -E_{x'}\tau_{2n} \tag{1b}$$

$$v_{z'n} = E_{z'}\tau_{0n} \tag{1c}$$

where

$$\left.\begin{aligned} \tau_{1n} &= B^{-1}\omega_n\nu_n/(\omega_n^2 + \nu_n^2) \\ \tau_{2n} &= B^{-1}\omega_n^2/(\omega_n^2 + \nu_n^2) \\ \tau_{0n} &= B^{-1}\omega_n/\nu_n \end{aligned}\right\} \tag{2}$$

(after CHAPMAN and BARTELS, 1940).

In the more general case, where the electric field has a y' component as well, we therefore have (the magnetic field is still in the z' direction)

$$v_{x'n} = E_{x'}\tau_{1n} + E_{y'}\tau_{2n} \tag{3a}$$

$$v_{y'n} = -E_{x'}\tau_{2n} + E_{y'}\tau_{1n} \tag{3b}$$

$$v_{z'n} = E_{z'}\tau_{0n} \tag{3c}$$

We can write equations (3a), (3b), (3c) in tensor form

$$\mathbf{v}_n = T_n \mathbf{E} \tag{4}$$

where the tensor T_n has the components

$$\left.\begin{aligned} &T_{x'x'n} = \tau_{1n}, \quad T_{x'y'n} = \tau_{2n}, \quad T_{y'x'n} = -\tau_{2n}, \quad T_{y'y'n} = \tau_{1n} \\ &T_{z'z'n} = \tau_{0n}, \quad T_{x'z'n} = T_{y'z'n} = T_{z'x'n} = T_{z'y'n} = 0 \end{aligned}\right\} \tag{4a}$$

Over most of the earth the magnetic field is inclined and it is therefore more convenient to use a coordinate system x, y, z with the x axis pointing south, the y axis east and the z axis vertically upward. Fig. 1 shows the relative position of the (x, y, z) (x', y', z') systems in the northern hemisphere. The y and y' axes are taken to be identical for simplicity and the angle of inclination denoted χ.

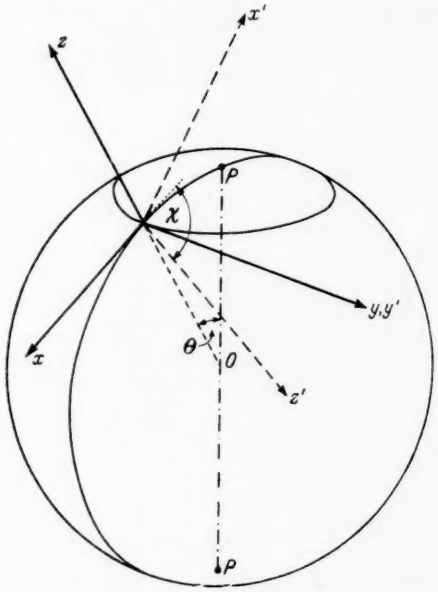


Fig. 1. Relative position of x, y, z and x', y', z' coordinate axes.

The direction cosines between the two systems are then given by $l_{xx'} = -\sin\chi$, $l_{xz'} = -\cos\chi$, $l_{yy'} = 1$, $l_{zx'} = \cos\chi$, $l_{zz'} = -\sin\chi$, $l_{xy'} = l_{yx'} = l_{yz'} = l_{z'y'} = 0$. and using these direction cosines the components of the tensor in the x, y, z system may be determined. They are given by

$$\left.\begin{aligned} &T_{xxn} = \tau_{1n}\sin^2\chi + \tau_{0n}\cos^2\chi,\quad T_{xyn} = -\tau_{2n}\sin\chi,\quad T_{xzn} = (\tau_{0n} - \tau_{1n})\sin\chi\cos\chi \\ &T_{yxn} = \tau_{2n}\sin\chi,\qquad\qquad T_{yyn} = \tau_{1n},\qquad\qquad T_{yzn} = -\tau_{2n}\cos\chi \\ &T_{zxn} = (\tau_{0n} - \tau_{1n})\sin\chi\cos\chi,\quad T_{zyn} = \tau_{2n}\cos\chi,\quad T_{zzn} = \tau_{1n}\cos^2\chi + \tau_{0n}\sin^2\chi \end{aligned}\right\} \quad (4b)$$

Equation (4) determines the drift velocity of the nth charged particle in the presence of known electric and magnetic fields.

If the velocities and number densities N_n of all charged particles are known the current density j_h at the height h can be determined. It is given by

$$j_h = \sum_n e_n N_n \boldsymbol{v}_n = \boldsymbol{E}\sum_n e_n N_n T_n = \boldsymbol{E}S \qquad (5)$$

VOL. 4 1953

The components of the conductivity tensor S are

$$\left.\begin{array}{lll} S_{xx} = \sigma_1 \sin^2\chi + \sigma_0 \cos^2\chi, & S_{xy} = -\sigma_2 \sin\chi, & S_{xz} = (\sigma_0 - \sigma_1) \sin\chi \cos\chi \\ S_{yx} = \sigma_2 \sin\chi, & S_{yy} = \sigma_1, & S_{yz} = -\sigma_2 \cos\chi \\ S_{zx} = (\sigma_0 - \sigma_1) \sin\chi \cos\chi, & S_{zy} = \sigma_2 \cos\chi, & S_{zz} = \sigma_1 \cos^2\chi + \sigma_0 \sin^2\chi \end{array}\right\} \quad (5a)$$

where

$$\left.\begin{aligned} \sigma_0 &= \sum_n e_n N_n \tau_{0n} = \sum_n \frac{e_n N_n}{B} \frac{\omega_n}{\nu_n} \\ \sigma_1 &= \sum_n e_n N_n \tau_{1n} = \sum_n \frac{e_n N_n}{B} \frac{\omega_n \nu_n}{\omega_n^2 + \nu_n^2} \\ \sigma_2 &= \sum_n e_n N_n \tau_{2n} = \sum_n \frac{e_n N_n}{B} \frac{\omega_n^2}{\omega_n^2 + \nu_n^2} \end{aligned}\right\} \quad (6)$$

If the ionosphere is regarded as a thin layer it is clear that the current density must be almost horizontal. A small vertical component may be present due to the fact that the current distribution with height may be different at different points of the earth. The current distribution varies, however, very slowly with position, except perhaps near the equator, and therefore the vertical current density is very much smaller than the horizontal current density.

If we therefore assume zero vertical current density we obtain from (5)

$$j_{zh} = S_{zx} E_x + S_{zy} E_y + S_{zz} E_z = 0$$

or

$$E_z = -E_x (S_{zx}/S_{zz}) - E_y (S_{zy}/S_{zz}) \quad (7)$$

Substituting (7) into the first two equations of (5) we obtain

$$j_{xh} = \sigma_{xxh} E_x + \sigma_{xyh} E_y \quad (8a)$$

$$j_{yh} = \sigma_{yxh} E_x + \sigma_{yyh} E_y \quad (8b)$$

where

$$\sigma_{xxh} = K^{-1} \sigma_1 \sigma_0 \quad (9a)$$

$$\sigma_{xyh} = -\sigma_{yxh} = -K^{-1} \sigma_2 \sigma_0 \sin\chi \quad (9b)$$

$$\sigma_{yyh} = K^{-1} (\sigma_1 \sigma_0 \sin^2\chi + \sigma_1 \sigma_3 \cos^2\chi) \quad (9c)$$

and

$$K = \sigma_1 \cos^2\chi + \sigma_0 \sin^2\chi \quad (9d)$$

$$\sigma_3 = \sigma_1 + \sigma_2^2/\sigma_1 \quad (9e)$$

Equations (8), (9) are also given by HIRONO (1952) and MARTYN (1952). MARTYN'S notation is adopted, with the exception that σ_2 differs in sign from the quantity, which MARTYN designates by the same symbol (σ_2 is negative here).

The conductivities σ_{xxh}, σ_{xyh}, σ_{yxh}, σ_{yyh} can be regarded as the components of a two-dimensional tensor which must be multiplied by the horizontal component of the electric field vector to give the current density. In this paper, however,

all equations will be given in terms of components. We shall refer to σ_{xxh}, σ_{yyh} as the direct conductivities and σ_{xyh} as the Hall conductivity.

If the current densities j_{xh}, j_{yh} are integrated over all the heights h of the ionosphere, we obtain

$$j_x = \sigma_{xx}E_x + \sigma_{xy}E_y \tag{10a}$$

$$j_y = \sigma_{yx}E_x + \sigma_{yy}E_y \tag{10b}$$

where

$$j_x = \int j_{xh}\,dh,\ j_y = \int j_{yh}\,dh,\ \sigma_{xx} = \int \sigma_{xxh}\,dh,\ \sigma_{xy} = -\sigma_{yx} = \int \sigma_{xyh}\,dh,\ \sigma_{yy} = \int \sigma_{yyh}\,dh$$

If the values of the height-integrated conductivities σ_{xx}, σ_{xy}, σ_{yx}, σ_{yy} and the wind velocity are known for every point of the earth, then it is possible to determine the differential equation for the (height-integrated) current density.

3. Differential Equations of the Dynamo Theory

The differential equations of the dynamo theory may be considerably simplified by assuming that the tidal winds are semidiurnal in character, that the conductivity of the ionosphere is independent of longitude and that the geographical and magnetic axes of the earth coincide. It is only after making these assumptions that the comparatively simple treatment to be described here becomes possible.

One further assumption not essential for the treatment is made here. The wind velocity is assumed independent of height in the conducting region. The significance of this is discussed in a later part of the paper.

There are two electric fields acting on the ions and electrons at any one point of the earth. If the reference system at the point concerned is stationary relative to the mean position of the air molecules, there is an induced electric field due to the vertical component of the earth's magnetic field and the horizontal movement of the tidal wind. In addition to this induced electric field there is also a polarization field, which is set up by accumulated charges in such a way that the current system set up by the combination of the two fields satisfies the equation of continuity.

Let the induced field $\boldsymbol{E}_i$ have the horizontal components

$$E_{ix} = E_{ixa}\sin 2\varphi + E_{ixb}\cos 2\varphi \tag{11a}$$

$$E_{iy} = E_{iya}\sin 2\varphi + E_{iyb}\cos 2\varphi \tag{11b}$$

where the coefficients E_{ixa} etc., are only functions of the colatitude θ and independent of the longitude angle φ (measured East of a reference meridian to be decided later) in accordance with our assumption about the semidiurnal nature of tidal velocities.

Let the potential, from which the polarization field can be derived, be given by

$$\psi = -r(\psi_a \sin 2\varphi + \psi_b \cos 2\varphi) \tag{12}$$

VOL. 4 1953

where r is the radius of the earth introduced here for convenience and ψ_a and ψ_b are again functions of only the colatitude θ and independent of the longitude angle φ. The horizontal components of the polarization field $\boldsymbol{E}_p = -\text{grad}\ \psi$ are

$$E_{px} = E_{pxa} \sin 2\varphi + E_{pxb} \cos 2\varphi = \frac{\partial \psi_a}{\partial \theta} \sin 2\varphi + \frac{\partial \psi_b}{\partial \theta} \cos 2\varphi \tag{13a}$$

$$E_{py} = E_{pya} \sin 2\varphi + E_{pyb} \cos 2\varphi = -\frac{2\psi_b}{\sin \theta} \sin 2\varphi + \frac{2\psi_a}{\sin \theta} \cos 2\varphi \tag{13b}$$

The current density $\boldsymbol{j}$ at any one point is independent of the reference system, if the total charge density can be neglected, and is determined by the sum of the induced and the polarization fields $\boldsymbol{E} = \boldsymbol{E}_i + \boldsymbol{E}_p$ and the height integrated conductivities σ_{xx}, σ_{xy}, σ_{yx}, σ_{yy} if it is assumed that both $\boldsymbol{E}_i$ and $\boldsymbol{E}_p$ are independent of height. It is further assumed that $E_{ixb} = E_{iya} = 0$ (it will be seen later that this is in accordance with present knowledge of tidal winds and at the same time determines the reference meridian). Applying (10), (12) and (13) gives

$$j_x = \left[\sigma_{xx}\left(E_{ixa} + \frac{\partial \psi_a}{\partial \theta}\right) - \sigma_{xy} \frac{2\psi_b}{\sin \theta}\right] \sin 2\varphi + \left[\sigma_{xx} \frac{\partial \psi_b}{\partial \theta} + \sigma_{xy}\left(E_{iyb} + \frac{2\psi_a}{\sin \theta}\right)\right] \cos 2\varphi \tag{14a}$$

$$j_y = \left[-\sigma_{xy}\left(E_{ixa} + \frac{\partial \psi_a}{\partial \theta}\right) - \sigma_{yy} \frac{2\psi_b}{\sin \theta}\right] \sin 2\varphi + \left[-\sigma_{xy} \frac{\partial \psi_b}{\partial \theta} + \sigma_{yy}\left(E_{iyb} + \frac{2\psi_a}{\sin \theta}\right)\right] \cos 2\varphi \tag{14b}$$

Substituting these values into the equation of continuity

$$\frac{\partial}{\partial \theta}(j_x \sin \theta) + \frac{\partial j_y}{\partial \varphi} = 0 \tag{15}$$

and equating the coefficients of $\sin 2\varphi$ and $\cos 2\varphi$ to zero gives the following two differential equations for ψ_a and ψ_b

$$\sigma_{xx} \sin \theta \frac{\partial^2 \psi_a}{\partial \theta^2} + \frac{\partial}{\partial \theta}(\sigma_{xx} \sin \theta) \frac{\partial \psi_a}{\partial \theta} - \frac{4\sigma_{yy}}{\sin \theta} \psi_a - 2 \frac{\partial \sigma_{xy}}{\partial \theta} \psi_b + \frac{\partial}{\partial \theta}(E_{ixa} \sigma_{xx} \sin \theta) - 2E_{iyb} \sigma_{yy} = 0 \tag{16a}$$

$$\sigma_{xx} \sin \theta \frac{\partial^2 \psi_b}{\partial \theta^2} + \frac{\partial}{\partial \theta}(\sigma_{xx} \sin \theta) \frac{\partial \psi_b}{\partial \theta} - \frac{4\sigma_{yy}}{\sin \theta} \psi_b + 2 \frac{\partial \sigma_{xy}}{\partial \theta} \psi_a + \frac{\partial}{\partial \theta}(E_{iyb} \sigma_{xy} \sin \theta) - 2E_{ixa} \sigma_{xy} = 0 \tag{16b}$$

Equations (16a), (16b) may be written in the form

$$\frac{\partial^2 \psi_a}{\partial \theta^2} + \alpha \frac{\partial \psi_a}{\partial \theta} - \beta \psi_a - \gamma \psi_b - \delta = 0 \tag{17a}$$

$$\frac{\partial^2 \psi_b}{\partial \theta^2} + \alpha \frac{\partial \psi_b}{\partial \theta} - \beta \psi_b + \gamma \psi_a - \varepsilon = 0 \tag{17b}$$

where the coefficients are functions of θ given by

$$\alpha = (\sigma_{xx} \sin\theta)^{-1} \frac{\partial}{\partial\theta}(\sigma_{xx} \sin\theta), \quad \beta = \frac{4\sigma_{yy}}{\sigma_{xx} \sin^2\theta}, \quad \gamma = \frac{2}{\sigma_{xx} \sin\theta} \frac{\partial\sigma_{xy}}{\partial\theta} \tag{18}$$

$$\delta = (\sigma_{xx} \sin\theta)^{-1} \left[\frac{\partial}{\partial\theta}(E_{ixa}\sigma_{xx} \sin\theta) - 2E_{iyb}\sigma_{yy}\right],$$

$$\varepsilon = (\sigma_{xx} \sin\theta)^{-1} \left[\frac{\partial}{\partial\theta}(E_{iyb}\sigma_{xy} \sin\theta) - 2E_{ixa}\sigma_{xy}\right]$$

Equations (17a) and (17b) form a system of two simultaneous linear differential equations for ψ_a and ψ_b, where the coefficients are functions of the independent variable. Such a system can be solved numerically in a comparatively simple manner (H. and B. S. JEFFREYS, 1950), as will be shown in section 4. If ψ_a and ψ_b are known, j_x and j_y can be determined from (14).

If the wind velocity and therefore E_i is a function of height (17a) and (17b) can be shown to be still valid but δ is given by

$$\delta = (\sigma_{xx} \sin\theta)^{-1} \left\{\int \left[\frac{\partial}{\partial\theta}(E_{ixa}\sigma_{xxh} \sin\theta) - 2E_{iyb}\sigma_{yyh}\right] \mathrm{d}h\right\}$$

and there is a similar expression for ε.

Before describing the numerical method used in this paper, alternative methods of obtaining solutions of these differential equations used by other authors will be shortly described.

MARTYN (1952) uses the fact that if a two-dimensional medium has direct conductivity σ_1, and Hall conductivity σ_2, and both are independent of the two coordinates, then the differential equation determining the current function from the electric field distribution is identical with the differential equation for an isotropic medium of conductivity $\sigma = \sigma_1 + \sigma_2^2/\sigma_1$.

This means that if $\sigma_2 \gg \sigma_1$, the "effective" conductivity σ is very much greater than either the direct or the Hall conductivity. The assumption of a conductivity independent of position is equivalent to the omission of all terms of the form $\partial\sigma/\partial\theta$ in (16).

The significance of this omission is best seen by setting $\sigma_{xx} = \sigma_{yy} = 0$. Then (16) reduces to

$$\psi_a = \left[2\frac{\partial\sigma_{xy}}{\partial\theta}\right]^{-1} \left[2E_{ixa}\sigma_{xy} - \frac{\partial}{\partial\theta}(E_{iyb}\sigma_{xy} \sin\theta)\right], \qquad \psi_b = 0$$

If $\partial\sigma_{xy}/\partial\theta = 0$ then ψ_a becomes infinite. If, as in practice, $\partial\sigma_{xy}/\partial\theta$ is of the same order as σ_{xy} then the polarization field derived from ψ_a will be of the same order as the induced field and the "effective conductivity" will only be slightly higher than σ_{xy}. The current distribution obtained by assuming an isotropic equivalent conductivity is therefore likely to be incorrect and the current densities obtained far too large, if the Hall conductivity is much greater than the direct conductivity. This will be further demonstrated by the results of the calculation to be described.

VOL. 4 1953

HIRONO (1952) by neglecting two terms of his equation (28) makes essentially similar assumptions to MARTYN and his work will therefore not be discussed here separately.

4. NUMERICAL METHOD OF SOLUTION

In the previous paragraph two simultaneous differential equations (17) were obtained for the potential functions ψ_a and ψ_b which may be written as

$$\frac{\partial^2\psi_a}{\partial\theta^2} = -\alpha\frac{\partial\psi_a}{\partial\theta} + \beta\psi_a + \gamma\psi_b + \delta \tag{19a}$$

$$\frac{\partial^2\psi_b}{\partial\theta^2} = -\alpha\frac{\partial\psi_b}{\partial\theta} + \beta\psi_b - \gamma\psi_a + \varepsilon \tag{19b}$$

For any given distributions of conductivities and tidal velocities with latitude and height, the coefficients α, β, γ, δ, ε can be computed as functions of the colatitude θ.

The boundary conditions satisfied by ψ_a and ψ_b are: ψ_a and ψ_b must both vanish at the poles (for $\theta = 0$ and $\theta = \pi$), $\partial\psi_a/\partial\theta$ and $\partial\psi_b/\partial\theta$ must both vanish at the equator (for $\theta = \pi/2$).

Assume that a method of obtaining solutions of (19) satisfying the boundary conditions at the equator is available. Let (ψ_{a1}, ψ_{b1}), (ψ_{a2}, ψ_{b2}) be two independent solutions obtained by setting $\delta = \varepsilon = 0$ and let ψ_{a3}, ψ_{b3} be a solution of the full equation.

Then $A\psi_{a1} + B\psi_{a2} + \psi_{a3}$, $A\psi_{b1} + B\psi_{b2} + \psi_{b3}$ is also a solution where A and B are constants. These are chosen to satisfy the boundary conditions at the pole by setting

$$A\psi_{a1}(0) + B\psi_{a2}(0) + \psi_{a3}(0) = 0 \tag{20a}$$

$$A\psi_{b1}(0) + B\psi_{b2}(0) + \psi_{b3}(0) = 0 \tag{20b}$$

and determining A and B from these equations.

In practice the solution ψ_a, ψ_b become infinite at the pole and it is sufficient to determine A and B from values ψ_a, ψ_b for a very small value of the colatitude θ.

Solutions of (19) are obtained in the following manner. An arbitrary set of values $\psi_a(\pi/2)$, $\psi_b(\pi/2)$ is selected for the equator. Knowing that $\partial\psi_a/\partial\theta$ and $\partial\psi_b/\partial\theta$ vanish, equations (19) determine $\partial^2\psi_a/\partial\theta^2$, $\partial^2\psi_b/\partial\theta^2$ at the equator. A Taylor expansion is then used up to the third term to determine ψ_a, ψ_b, $\partial\psi_a/\partial\theta$, $\partial\psi_b/\partial\theta$ at a latitude just north of the equator. If (19) is differentiated further, higher derivatives than the second of ψ_a and ψ_b can be determined and more than three terms of the Taylor series can be used. (This has not been done in the present calculation.)

The same procedure is now repeated at the new latitude and values of ψ_a, ψ_b, $\partial\psi_a/\partial\theta$, $\partial\psi_b/\partial\theta$ at a point still further north are obtained. Thus point by point the complete solution of the differential equation is determined. The smaller the interval between neighbouring points and the more terms used, the greater the accuracy obtained.

Having determined the potential ψ and its first derivative as functions of the latitude, the polarization field is determined from (13) and the (height integrated) current density from (14).

5. Atmospheric Tides and Induced Electric Field

The semidiurnal component of atmospheric pressure variation on the ground resulting from the solar tide is assumed to be given by the first term of SIMPSON's (1928) formula

$$p = 0{\cdot}937 \sin^3\theta \sin[2(\omega t + \lambda) + 154°] \tag{21}$$

in millimetres of mercury, where t is Greenwich mean time, ω the angular velocity of the earth and λ the longitude East of Greenwich.

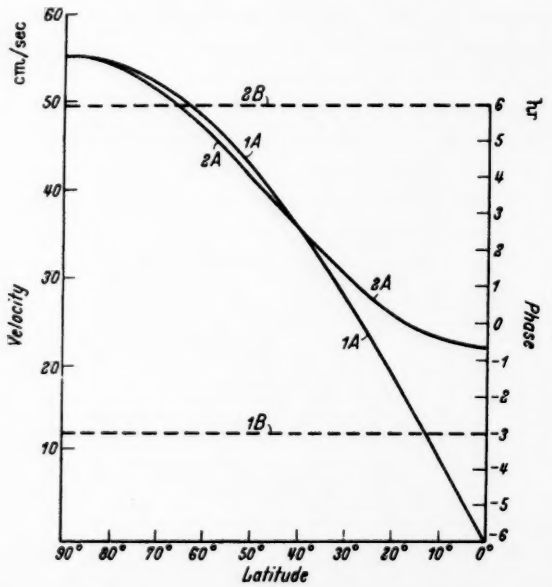


Fig. 2. Amplitudes and phases of the semidiurnal wind velocity components on the ground as functions of the latitude. Curve 1A shows the amplitude, curve 1B the phase of the South component v_{x0}. Curve 2A shows the amplitude, curve 2B the phase of the East component v_{y0}.

Writing $\omega t + \lambda + 77° = \varphi$ the pressure variation is given by $p = 0{\cdot}937 \sin^3\theta \sin 2\varphi$. The South and East components of the velocity of the air on the ground were derived from this pressure variation by BARTELS (1928). The numerical result may be expressed as

$$v_{x0} = 55{\cdot}2 \cos\theta \cos 2\varphi \text{ cm/sec} \tag{22a}$$

$$v_{y0} = -55{\cdot}2(0{\cdot}7 + 0{\cdot}3 \cos 2\theta) \sin 2\varphi \text{ cm/sec} \tag{22b}$$

These wind velocity components are shown by Fig. 2 as functions of latitude.

VOL. 4 1953

From (22) the induced electric field can be determined. Assuming for simplicity that the earth's geographic and magnetic axes coincide, the result is

$$E_{ix} = E_{ixa} \sin 2\varphi, \qquad E_{iy} = E_{iyb} \cos 2\varphi \tag{23a}$$

where

$$\left.\begin{aligned} E_{ixa} &= 3{\cdot}46.10^{-5} \cos\theta(0{\cdot}7 + 0{\cdot}3 \cos 2\theta) \text{ volts/metre} \\ E_{iyb} &= 3{\cdot}46.10^{-5} \cos^2\theta \text{ volts/metre} \end{aligned}\right\} \tag{23b}$$

Equations (22), (23) represent wind velocities and induced electric fields on the ground. In the ionosphere (TAYLOR, 1936; PEKERIS, 1937; WEEKES and WILKES. 1947) these values are multiplied by an unknown multiplication factor and, depending on the exact temperature distribution, they may be either in phase equality with or in phase opposition to those on the ground.

In the calculation the values on the ground have been used. The tidal amplification is then estimated by the discrepancy between calculated and "measured" values of current density.

All results will be obtained relative to the phase of the atmospheric pressure tide. Maximum pressure occurs about two hours before noon or midnight for the solar tide (SIMPSON, 1928) and about an hour after lunar transit for the lunar tide. (More details of the atmospheric lunar tides are given by CHAPMAN and TSCHU 1948.)

6. The Height Integrated Conductivities

A simplified method for the calculation of σ_{xx}, σ_{xy} and σ_{yy} is used here. This method is not quite accurate but it avoids carrying out numerical integrations of σ_{xxh}, σ_{xyh}, σ_{yyh} for each latitude.

Three similar equations to (9) are used:

$$\sigma_{xx} = L^{-1}[\int\sigma_1 \, dh]\,[\int\sigma_0 \, dh] \tag{24a}$$

$$\sigma_{xy} = -L^{-1}[\int\sigma_2 \, dh]\,[\int\sigma_0 \, dh] \tag{24b}$$

$$\sigma_{yy} = L^{-1}\{[\int\sigma_1 \, dh]\,[\int\sigma_0 \, dh] \sin^2\chi + [\int\sigma_1 \, dh]\,[\int\sigma_3 \, dh] \cos^2\chi\} \tag{24c}$$

$$L = [\int\sigma_1 \, dh] \cos^2\chi + [\int\sigma_0 \, dh] \sin^2\chi \tag{24d}$$

These equations are very nearly correct at the equator and at points further than about 10° from the equator. If a value of $\int\sigma_0 \, dh$ lower than the actual one is chosen. the transition from the equator to a latitude 10° off the equator can be fairly well approximated in σ_{xy} and σ_{yy} at the expense of having a somewhat low value of σ_{xx} at the equator. This is not very important because the electric field has no South component at the equator.

The following values (in electromagnetic units) were chosen to represent midday conditions at sunspot maximum:

$$\left.\begin{aligned} \int\sigma_0 \, dh &= 2{\cdot}10^{-7}(1 + 2\sin\theta) \\ \int\sigma_1 \, dh &= 2{\cdot}10^{-9}(1 + 2\sin\theta) \\ \int\sigma_2 \, dh &= 7{\cdot}10^{-9}(1 + 2\sin\theta) \\ \int\sigma_3 \, dh &= 6{\cdot}10^{-8}(1 + 2\sin\theta) \end{aligned}\right\} \tag{25}$$

The factor $(1 + 2 \sin \theta)$ was introduced to take the variations of electron density with latitude into account. The last three numerical factors have about half the values given by MARTYN (1952) for the height integrated conductivities of the ionosphere. MARTYN also gives more detailed information on the variation of the conductivities σ_1, σ_2, σ_3 with height and on the variation of σ_{xxh}, σ_{xyh}, σ_{yyh} at a constant height with latitude.

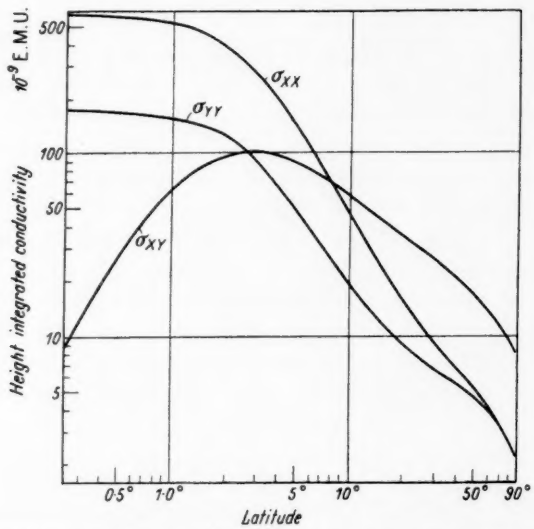


Fig. 3. The height-integrated conductivities σ_{xx}, σ_{xy} and σ_{yy} as functions of the latitude.

The values of σ_{xx}, σ_{xy} and σ_{yy} as functions of latitude computed with the aid of assumptions (23), (24) are shown in Fig. 3. On approaching the equator σ_{yy} rises to a large maximum and σ_{xy} rapidly decreases to zero over a few degrees of latitude.

7. CALCULATED CURRENT SYSTEM

Before describing the results of the calculation some details of the numerical work are given.

The conductivities σ_{xx}, σ_{xy}, σ_{yy} were computed to five figure accuracy (using a *Facit E.S.A.* calculating machine). The coefficients α, β, γ, δ, ε were then calculated to the same nominal accuracy. Some of the differentiations were, however, carried out numerically; while this reduced the accuracy, it saved time.

The coefficients were calculated for values of θ at $\frac{1}{4}°$ interval between 1°–5°, at $\frac{1}{2}°$ intervals between 5°–10° and 80°–90°, at 1° intervals between 10°–20° and 70°–80° and at 2° intervals between 20°–70°. These coefficients are shown by Fig. 4. It can be seen that the coefficients tend to infinity near the pole. The three solutions are therefore carried as far as $\theta = 1°$, the accuracy being increased to 6 figures over the last ten degrees.

The first two solutions involve identical numerical work if the values $\psi_{a1}(\pi/2) = 1$, $\psi_{b1}(\pi/2) = 0$, $\psi_{a2}(\pi/2) = 0$, $\psi_{b2}(\pi/2) = 1$ are chosen, so that the

VOL. 4 1953

differential equation requires to be solved once to obtain both solutions. For the last solution $\psi_{a3}(\pi/2) = 0$, $\psi_{b3}(\pi/2) = 0$ was chosen. The solutions were carried up to $\theta = 1°$, where equation (20) was applied. Over the interval 10°–1° the solutions were calculated up to six figures, because in the final superposition of the three solutions differences between large and almost equal numbers occur.

The solutions in terms of ψ_a and ψ_b are shown by Fig. 5. From these the current densities were calculated. These are given for a tidal amplification of

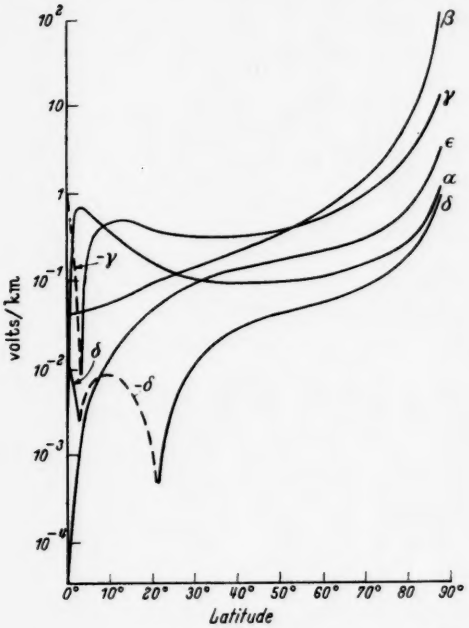


Fig. 4. The coefficients α, β, γ, δ, ε of the differential equations (19) as functions of the latitude. When one of the coefficients is negative, the absolute value is indicated by a dotted curve.

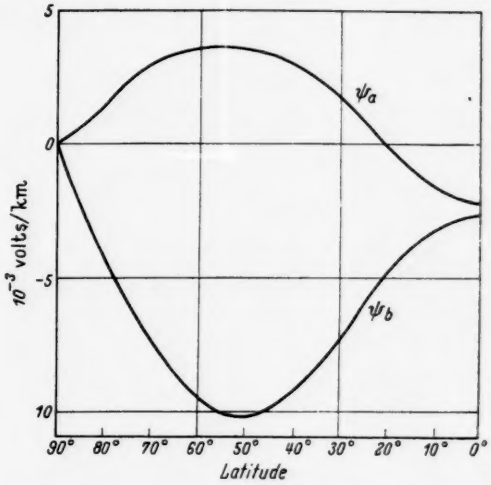


Fig. 5. The solutions ψ_a, ψ_b of the differential equations (19) as functions of the latitude.

unity by Fig. 6. The calculated polarization fields, induced fields and total fields are shown by Fig. 7, Fig. 8 and Fig. 9. Figs. 6, 7, 8 and 9 show the maximum values of East and South components of current density and electric field and also the number of hours by which the occurrences of these maxima lag behind maximum atmospheric pressure.

If the ionospheric air tide differs in amplitude by a known factor and in phase by a known number of hours from the ground tide, then the amplitudes given by the preceding Figures must be multiplied by the tidal amplification, and the phase difference between ionospheric and ground air tides must be added to the phases shown there.

The current densities displayed by Fig. 6 may now be compared with the current distribution calculated by BARTELS (1928) from CHAPMAN'S (1919) harmonic analysis of the magnetic variations. BARTELS' Fig. 28 gives a maximum

current density of about 18 amp/km at high latitudes for 1902, for the equinoxes and for the most intense current circuit which is almost centred on noon. The

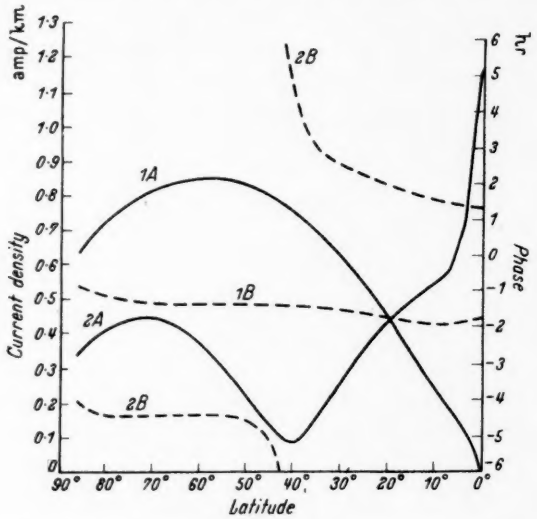


Fig. 6. The amplitudes and phases of the semidiurnal current density components as functions of the latitude. Curve 1A shows the amplitude, curve 1B the phase of the South component j_x. Curve 2A shows the amplitude, curve 2B the phase of the East component j_y.

VOL. 4 1953

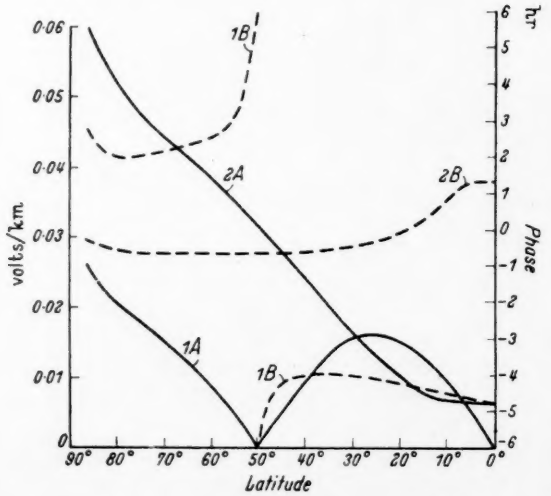


Fig. 7. The amplitudes and phases of the electric polarization field components as functions of the latitude. Curve 1A shows the amplitude, curve 1B the phase of the South component E_{px}. Curve 2A shows the amplitude, curve 2B the phase of the East component E_{py}.

current density for sunspot maximum (1905) is higher by $\frac{1}{3}$ according to BARTELS (1928); this gives 24 amp/km. If noon values of conductivity were maintained, as our calculation assumes, during the whole day, then this current density would

probably increase to about 30 amp/km and, if we compare this with 0·45 amp/km shown by Fig. 6, we are led to a tidal amplification of about 65.

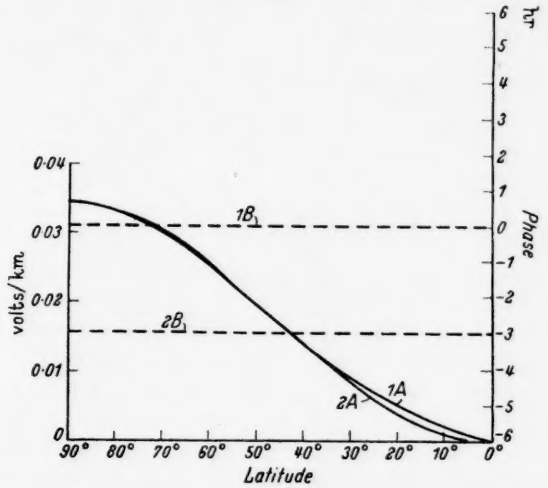


Fig. 8. The amplitudes and phases of the induced electric field components E_{ix}, E_{iy} as functions of the latitude. Notation as in Fig. 7.

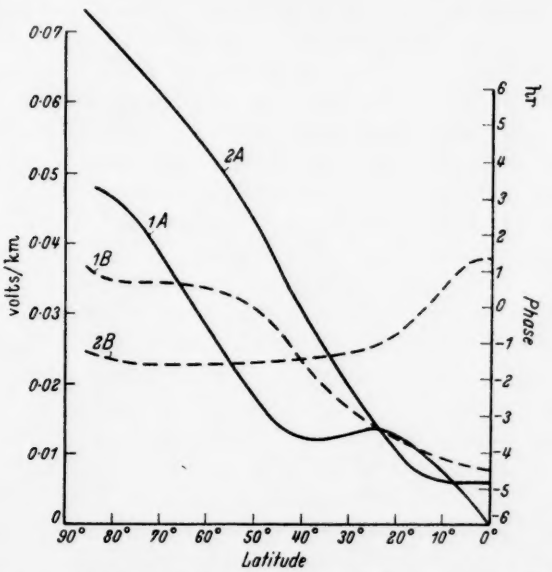


Fig. 9. The amplitudes and phases of the total electric field components E_x, E_y as functions of the latitude. Notation as in Fig. 7.

Consideration of the total current in the circuit leads to the same result. Fig. 6 shows a total current of about 1800 amps. as calculated by integration. BARTELS (1928) gives 62,000 amperes for 1902, which can again be modified to 83,000 amperes for sunspot maximum and to 104,000 amperes if the noon conductivity

is assumed to exist at all times. This leads to a tidal amplification of 58. A tidal amplification of about 60 is therefore a reasonable estimate. It agrees fairly well with measured "wind" velocities (PHILLIPS, 1952).

The tidal amplification factor of 60 obtained here is much lower than the usual estimates, even than the estimate of MARTYN (1952) whose "effective conductivity" would tend to rather underestimate this factor. The methods used by other authors to determine tidal amplification are therefore shortly outlined here.

CHAPMAN (1919) found that the height integrated conductivity of the ionosphere (assumed isotropic) was $2{\cdot}5.10^{-5}$ e.m.u. if the winds at ionospheric heights were the same as at ground levels. CHAPMAN in his work did not use the tidal wind distribution of BARTELS given by equation (22). His wind velocity which was derived from the atmospheric pressure variations by neglecting the Coriolis force, is identical with that of BARTELS (1928) at the equator, but it is $4\frac{1}{4}$ times lower at a latitude of 60° and drops to zero at the pole. This shows that CHAPMAN would have obtained a considerably lower conductivity than $2{\cdot}5.10^{-5}$ e.m.u. had he taken the Coriolis force into account. Using this lower value of conductivity, lower values of tidal amplification would have been obtained by previous authors who, although they usually accepted the velocity distribution of BARTELS (see Fig. 12 in Special Report No. 2 of U.R.S.I.),* used the high conductivity based on a different velocity distribution.

There are other interesting properties of the current distribution displayed by Fig. 6. It will be noticed that the phase of the East component of the induced electric field and the phase of the East component of the current density at high latitudes are about $1\frac{1}{2}$ hours out of phase. On MARTYN'S assumption of an equivalent scalar conductivity, there should be no phase difference. This again shows that terms of the type $\partial\sigma/\partial\theta$ change the character of the differential equation (16).

According to the current distribution of Fig. 6, maximum Eastern current should occur about $4\frac{1}{2}$ hours before pressure maximum at high latitudes. Pressure maximum occurs at 10 hrs (after midday or midnight; this notation is used throughout the paper) for the solar atmospheric tide on the ground, while BARTELS' (1928) current distribution shows maximum Western current density at about 11 hrs corresponding to maximum Eastern current density at about 5 hrs. This would seem to indicate that the tidal wind system causing the dynamo current is almost exactly in phase with the wind system on the ground.

Considering lunar tides the time of maximum pressure on the ground varies with the seasons (CHAPMAN and TSCHU, 1946) a reasonable mean value being 1 hr while maximum Eastern current density occurs according to BARTELS' (1928) lunar current distribution at about 3 hrs at high latitudes. This seems to show that the lunar air tide determining currents in the ionosphere is almost exactly in phase opposition to the lunar air tide on the ground.

It is rather interesting that ground and ionospheric air tides are in phase for the solar but out of phase for the lunar tide. Differences between the distribution of phases and amplitudes with height of these two tides may be due to the larger amplitude of the solar tide (WILKES, 1952).

---

* MARTYN, D. F.; U.R.S.I. Special report on tidal phenomena in the ionosphere.

VOL. 4 1953

It is further seen from Fig. 6 that the current density increases by a factor greater than two over the last ten degrees of latitude when approaching the equator. The current density at the equator is more than three times as high as the maximum current density at high latitudes. This explains the abnormally large magnetic variations near the magnetic equator.

8. Vertical Electron Drifts

It was shown in section 2 that the vertical current density in the ionosphere must be extremely small. Large vertical electron drifts are, however, possible if they

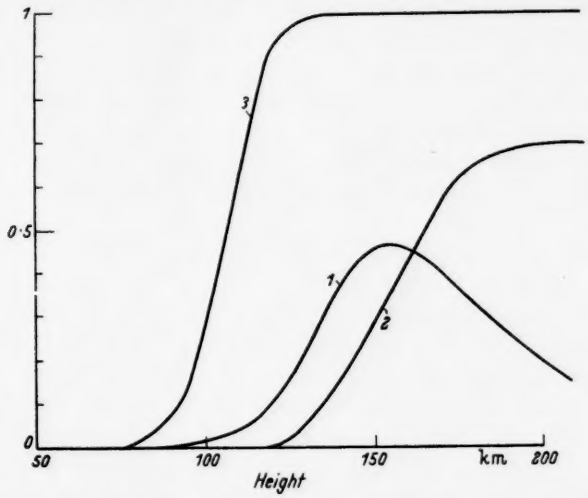


Fig. 10. The dimensionless coefficients BC_1, BC_2 in the equations for vertical electron drift velocity components. Curve 1 represents $-BC_1$ at 45° latitude, curve 2 shows BC_2 at 45° latitude and curve 3 indicates BC_2 at the equator as functions of height.

are accompanied by a drift of positive ions in the same direction. This electron drift is readily calculated from the results in sections 4 and 6.

The vertical electron velocity expressed in terms of the tensor T of equation (4) (the subscript e stands for electron) is given by

$$v_{ze} = T_{zxe}E_x + T_{zye}E_y + T_{zze}E_z \tag{26}$$

Substituting E_z from (7) combined with (5a) and T_{zxe}, T_{zye}, T_{zze} from (4b) gives

$$v_{ze} = C_1E_x + C_2E_y \tag{27a}$$

where

$$\left.\begin{aligned} C_1 &= (\tau_{0e} - \tau_{1e}) \sin\chi\cos\chi - (\tau_{1e}\cos^2\chi + \tau_{0e}\sin^2\chi)\frac{(\sigma_0 - \sigma_1)\sin\chi\cos\chi}{\sigma_1\cos^2\chi + \sigma_0\sin^2\chi} \\ \text{and} \\ C_2 &= \tau_{2e}\cos\chi - (\tau_{1e}\cos^2\chi + \tau_{0e}\sin^2\chi)\frac{\sigma_2\cos\chi}{\sigma_1\cos^2\chi + \sigma_0\sin^2\chi} \end{aligned}\right\} \tag{27b}$$

The coefficients C_1, C_2 have been calculated using the same assumption about atmospheric properties, as Martyn used in the calculation of σ_1, σ_2, σ_3. Fig. 10

shows $-BC_1$ and BC_2 as functions of height for a latitude of 45° as well as BC_2 for the equator. (C_1 vanishes at the equator.) It is readily seen that below a height of about 170 km C_1, which is negative, predominates while above 170 km C_2 predominates. Thus the phase of upward electron drift is determined by E_x in the E-layer but by E_y in the F_1- and F_2-layers. There is some uncertainty about winds in the F_1- and F_2-regions which in any case contribute little to the dynamo current system. Here we shall assume that there are no tides in these regions and therefore $E_y = E_{py}$, $E_{iy} = 0$.

We thus conclude from Fig. 9 that, in the E-layer maximum upward electron drift occurs 5–6 hours before maximum pressure at high latitudes and about 1·5 to 3 hours after maximum pressure at low latitudes. For the lunar tide this means 1–2 hrs at high latitudes and 8–10 hrs at low latitudes. The lunar tide in h_E' takes place at about 11·3 hrs lunar time at high latitudes (APPLETON and WEEKES, 1939) and at about 6 hrs at low latitudes (MARTYN, 1948), in both cases earlier than the theory predicts. This is not a serious discrepancy in view of the assumed over-simplified distribution of the phase and amplitude of tidal velocity with height.

In the F_1- and F_2-layers Fig. 7 shows that maximum upward electron drift occurs at about an hour before maximum air pressure at latitudes higher than about 30° and about an hour after maximum air pressure at the equator. For the lunar tide this means 6 hrs for latitudes higher than 30° and 8 hrs for the equator. This agrees well with measured values (MARTYN, 1947, 1948).

It should be noted here that the existence of a phase reversal in the lunar tide in h_E between high and low latitudes and the lack of such a phase reversal in h_{1F}' and h_{2F}' is explained by the fact that the former is determined by the North component, whereas the latter two are determined by the East component of the electric field. The non-existence of the phase reversal in the F-layer therefore does not indicate that there is no atmospheric tide in the F-layer as was formerly believed. Conversely the existence of the phase reversal in the E-layer does not prove that there is an atmospheric tide there.

For the solar tide we obtain 9 hrs at latitudes higher than 30° and 11 hrs at the equator as times of maximum upward electron drift. A. P. MITRA (1950) shows that the average variation of h_{2F}' at Delhi is almost semidiurnal in character, a second smaller maximum being reached at about 10 p.m. after a first maximum at about midday.

Lunar effects in absorption were detected by APPLETON and BEYNON (1949). It appears that maximum downward drift occurs at 10·9 lunar hours. This would appear to indicate that the lunar atmospheric pressure variations in the D- and E-layers have opposing phases.

9. HORIZONTAL ELECTRON DRIFTS

In recent years many measurements (PHILLIPS, 1952; SALZBERG and GREENSTONE, 1951) of ionospheric "winds" have been carried out by recording the fading of a single ionospheric echo with two spaced receivers. This method may measure the movements of electrons and not of air. It is therefore of interest to determine the horizontal electron drift velocities resulting from the theory.

VOL. 4 1953

The x and y components of electron drift velocity relative to the ground are given by

$$v_{xe} = v_{x0} + E_x T_{xxe} + E_y T_{xye} + E_z T_{xze}$$
$$v_{ye} = v_{y0} + E_x T_{yxe} + E_y T_{yye} + E_z T_{yze}$$

where v_{x0}, v_{y0} are the horizontal components of the air velocity. Substitutions of E_z and T_e as in the previous paragraph give

$$v_{xe} = v_{x0} + E_x\left[\tau_{1e}\sin^2\chi + \tau_{0e}\cos^2\chi - (\tau_{0e} - \tau_{1e})\sin\chi\cos\chi\,\frac{(\sigma_0 - \sigma_1)\sin\chi\cos\chi}{\sigma_1\cos^2\chi + \sigma_0\sin^2\chi}\right]$$
$$+ E_y\left[-\tau_{2e}\sin\chi - (\tau_{0e} - \tau_{1e})\sin\chi\cos\chi\,\frac{\sigma_2\cos\chi}{\sigma_1\cos^2\chi + \sigma_0\sin^2\chi}\right] \qquad (28a)$$

$$v_{ye} = v_{y0} + E_x\left[\tau_{2e}\sin\chi + \tau_{2e}\cos\chi\,\frac{(\sigma_0 - \sigma_1)\sin\chi\cos\chi}{\sigma_1\cos^2\chi + \sigma_0\sin^2\chi}\right]$$
$$+ E_y\left[\tau_{1e} + \tau_{2e}\cos\chi\,\frac{\sigma_2\cos\chi}{\sigma_1\cos^2\chi + \sigma_0\sin^2\chi}\right] \qquad (28b)$$

In most parts of the ionosphere not too near the equator, these equations are well approximated by

$$v_{xe} = v_{x0} + E_y\left[-\tau_{2e}\sin\chi - \frac{\tau_{0e}\sigma_2}{\sigma_0}\frac{\cos^2\chi}{\sin\chi}\right] \qquad (29a)$$

$$v_{ye} = v_{y0} + E_x\tau_{2e}/\sin\chi \qquad (29b)$$

These equations can be further simplified in the E-region to

$$v_{xe} = v_{x0} - E_y\tau_{2e}/\sin\chi \qquad (30a)$$

$$v_{ye} = v_{y0} + E_x\tau_{2e}/\sin\chi \qquad (30b)$$

while in F_1- and F_2-regions assuming $v_{x0} = v_{y0} = 0$

$$v_{xe} = -E_{py}\sin\chi/B \qquad (31a)$$

$$v_{ye} = E_{px}/(B\sin\chi) \qquad (31b)$$

In (30) v_{x0} and v_{y0} are equal and opposite in sign to the contributions by the induced voltages and thus in the E-region:

$$v_{xe} = -E_{py}/(B\sin\chi) \qquad (32a)$$

$$v_{ye} = E_{px}/(B\sin\chi) \qquad (32b)$$

The North component of electron drift both in the E- and F-regions is thus in phase with E_{py}, while the East component is in phase with E_{px}. Maximum North component thus occurs (Fig. 7) about one hour before maximum pressure over most of the earth, while maximum East component occurs two hours after maximum pressure at high latitudes and 4½ hours before maximum pressure at low latitudes.

Assuming, in accordance with previous conclusions, that the solar air tide is in phase with the tide on the ground, maximum North component should occur at 9 hrs, maximum East component at 12 hrs for high latitudes and at $5\frac{1}{2}$ hrs for low latitudes.

These values are in approximate phase opposition to those measured by PHILLIPS. It appears that the "winds" measured by PHILLIPS were not electron drifts but wave-like motions in the air of random direction which influenced the electron density and on which the semidiurnal air velocity was superimposed. Therefore the measured mean velocity was the velocity of the air and, assuming the solar tide to be in phase with the tide on the ground, agreement between theory and experiment is good.

Other measurement of winds in the ionosphere using meteor trails and the scintillation of radio stars have been reported. Measurements on meteor trails indicate the electron drift and are therefore of special interest. Preliminary measurements by GREENHOW (1952), which agree with similar measurements of MANNING, VILLARD and PETERSON (1950), indicate that the wind directions measured by them are opposite to those measured by PHILLIPS (1952) and SALZBERG and GREENSTONE (1951) and agree with the results of the theory presented here.

10. CONCLUSIONS

The numerical method given here leads to a current system which differs both in phase and in amplitude from that predicted by MARTYN'S (1952) method. Using this numerical method the tidal amplification is estimated as about 60.

The calculated phase of the current system indicates that the solar atmospheric tide in the *E*-region is in phase with the tide on the ground, but that the lunar air tide in *E*-region is in phase opposition to the ground tide.

The calculated vertical electron drifts fit approximately the known ionospheric lunar height tides. The calculated horizontal electron drifts agree with observations on long duration meteor echoes but the calculated phases are opposed to those obtained from fading measurements. The fading measurements appear therefore to refer to air movements and not to electron drifts.

Acknowledgements—This paper is published with the permission of the South African Council for Scientific and Industrial Research.

I wish to thank Dr. D. F. MARTYN, who has been kind enough to send a draft of his "Nature" article in advance of publication.

REFERENCES

APPLETON, E. V.	1937	Proc. Roy. Soc. A. **162,** 451
APPLETON, E. V. and BEYNON, W. J. G.	1948	Nature **162,** 486
APPLETON, E. V., and WEEKES, K.	1939	Proc. Roy. Soc. A. **171,** 171
BALFOUR STEWART	1882	Encyclopedia Britannica 9th Edition
BARTELS, J.	1928	Handbuch der Experimentalphysik 25
BATES, D. R., and MASSEY, H. S. W.	1951	J. Atmosph. Terr. Phys. **2,** 1
CHAPMAN, S.	1919	Phil. Trans. Roy. Soc. **218A,** 1
CHAPMAN, S., and BARTELS, J.	1940	Geomagnetism, Clarendon Press Oxford
CHAPMAN, S., and TSCHU, K. K.	1948	Proc. Roy. Soc. A. **195,** 310

VOL. 4 1953

COWLING, T. G.	1945	Proc. Roy. Soc. A. **183,** 453
GREENHOW, J. S.	1952	J. Atmosph. Terr. Phys. **2,** 282
HIRONO, M.	1952	Journal of Geomagnetism and Geoelectricity **4,** 7
JEFFREYS, H., and B. S.	1950	Mathematical Physics, p. 290, Cambridge University Press
MANNING, L. A., VILLARD, O. G., and PETERSON, A. M.	1950	Proc. Inst. Radio Engrs. **38,** 887
MARTYN, D. F.	1947	Proc. Roy. Soc. A. **190,** 273
	1948	Proc. Roy. Soc. A. **194,** 429
	1952	Nature **170,** 1090
MITRA, A. P.	1950	Ind. Phys. **24,** 387
PEKERIS, C. L.	1937	Proc. Roy. Soc. A. **158,** 650
PHILLIPS, G. J.	1952	J. Atmosph. Terr. Phys. **2,** 141
SALZBERG, C. D., and GREENSTONE, R.	1951	J. Geophys. Res. **56,** 521
SCHUSTER, A.	1908	Phil. Trans. Roy. Soc. A. **163,** 208
SIMPSON, J. C.	1928	Quart. J. Roy. Met. Soc. **44,** 1
TAYLOR, G. I.	1936	Proc. Roy. Soc. A. **156,** 318
WEEKES, K., and WILKES, M. V.	1947	Proc. Roy. Soc. A. **192,** 80
WILKES, M. V.	1952	Quart. J. Roy. Met. Soc. **78,** 321

OL.
4
953

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 204 to 218. Pergamon Press Ltd., London

The measurement of ionospheric absorption using observations of 18·3 Mc/s cosmic radio noise

A. P. MITRA*,† and C. A. SHAIN*

(*Received* 10 *September* 1953)

ABSTRACT

A new technique permits measurement of the total attenuation suffered by 18·3 Mc/s cosmic radio noise passing completely through the ionosphere. It is shown that the total absorption may be divided into two components, one due to absorption mainly in the D region and the other in the F_2 region. The observations of D region absorption confirm the diurnal and seasonal variations observed by other workers. F_2 absorption depends on the critical frequency but not on the height of the region and there is evidence of increased absorption at night which may be caused by irregularities in the upper F region. The observational results are compared with those obtained by other workers and suggestions are made for the use of the method, particularly in the exploration of the upper F_2 region.

1. INTRODUCTION

The usual method of measurement of attenuation of radio waves passing through the ionosphere is to compare the amplitudes of the multiple reflections of pulses on several frequencies at vertical incidence. Such measurements have been supplemented, at several stations, by measurement of the absorption, both of pulses and of continuous waves, at oblique incidence. In all cases the measurements are made on radio waves transmitted from the ground and reflected from the ionosphere, and great care is needed to obtain accurate results, chiefly owing to the irregular fluctuations in the amplitude of the received signals. A survey of the problem, with a summary of the more important results obtained at Slough. England, is given by SMITH-ROSE (1951).

In recent years some observations have been made of the attenuation of radio waves which have passed through the whole of the earth's atmosphere. Observations of moon echoes on frequencies near 20 Mc/s (KERR and SHAIN. 1951) have shown that useful new information can be obtained by this method. However, observations of moon echoes require large transmitters and, up to the present. have all been made at low angles of elevation. In one of the earliest observations of radiation, JANSKY (1937) noted that during the day the intensity of the cosmic noise was lower than expected and this was attributed to ionospheric absorption. In 1949, a preliminary series of observations of cosmic noise at 18·3 Mc/s was undertaken at the Radiophysics Laboratory (SHAIN, 1951) and it was apparent that, although the ionospheric attenuation was comparatively small, under certain conditions it was appreciable and could be measured without difficulty. These observations were made at vertical incidence.

This paper describes the results of an extensive study of ionospheric absorption using a series of more accurate observations of cosmic noise at 18·3 Mc/s which were taken after the series referred to above. The site, as before, was Hornsby (near Sydney, New South Wales, latitude 34°S., longitude 151°E.). The primary

* Division of Radiophysics, Commonwealth Scientific and Industrial Research Organization, Australia.

† On leave from the Institute of Radiophysics and Electronics, University of Calcutta. and the Council of Scientific and Industrial Research, Government of India.

VOL. 4 1953

purpose of these observations was again the study of cosmic and solar noise, and those aspects of the work will be described elsewhere (SHAIN and HIGGINS, 1954). However, it was thought that the recorded data might be used to make a detailed study of the effect of the ionosphere on the intensity of the incoming radiation as measured at the surface of the earth. As the work proceeded it was found that the small variations in intensity due to the ionosphere could readily be measured.

Attenuation as measured by the present method is the sum of contributions arising in all the different regions of the ionosphere. However, it is shown that the major contributions are due to the D and F_2 regions and a method of separating these contributions is developed.

2. TECHNIQUE

The equipment used for the cosmic noise observations has been described fully elsewhere (SHAIN and HIGGINS, 1954). Briefly, ordinary communications-type receivers were used in conjunction with an aerial consisting of an array of 30 half-wave dipoles so arranged that the aerial beam was normally directed upwards (although the actual direction could be varied) and had an overall beam-width to half-power of 17°. At the frequency used, 18·3 Mc/s, the intensity of cosmic noise is high and it appears that for measurements of ionospheric effects on cosmic noise, a frequency in the range 15–20 Mc/s is most suitable; at lower frequencies, interference from radio stations and from atmospherics becomes serious. At 18·3 Mc/s the period of the records lost from these causes was small and any such interference could be recognized from the characteristics of the record. Typical records are shown in Fig. 1.

For the present purpose, ionospheric absorption at any time is specified by the ratio, expressed in decibels, of P, the received cosmic noise power, to P_0, the power that would have been received if there had been no absorption at that time.

The variations in received power P consist of variations due to ionospheric absorption together with variations in P_0 as the aerial beam scans different regions of the sky. The latter variation is a function only of aerial direction and of sidereal time and could be eliminated if the values of P_0 for the whole sidereal day and for the aerial directions of interest could be determined.

The present method of measuring absorption involves taking observations of cosmic noise for about a year with a fixed aerial, so that all parts of the sky scanned by the aerial beam will have been observed at all hours of the day and night. From these observations it is possible to select those for which ionospheric attenuation is negligible. With these selected observations, a curve of the unabsorbed intensity of cosmic noise against sidereal time can be drawn and the absorption under any conditions can be calculated by comparing the observed intensity with the "standard" intensity for the sidereal time of the observation. If the aerial beam can be directed to cover different regions of the sky, a standard curve can be drawn from each position.*

* A similar, but slightly different, technique has been suggested by LAFFINEUR and KOURGANOFF (1950).

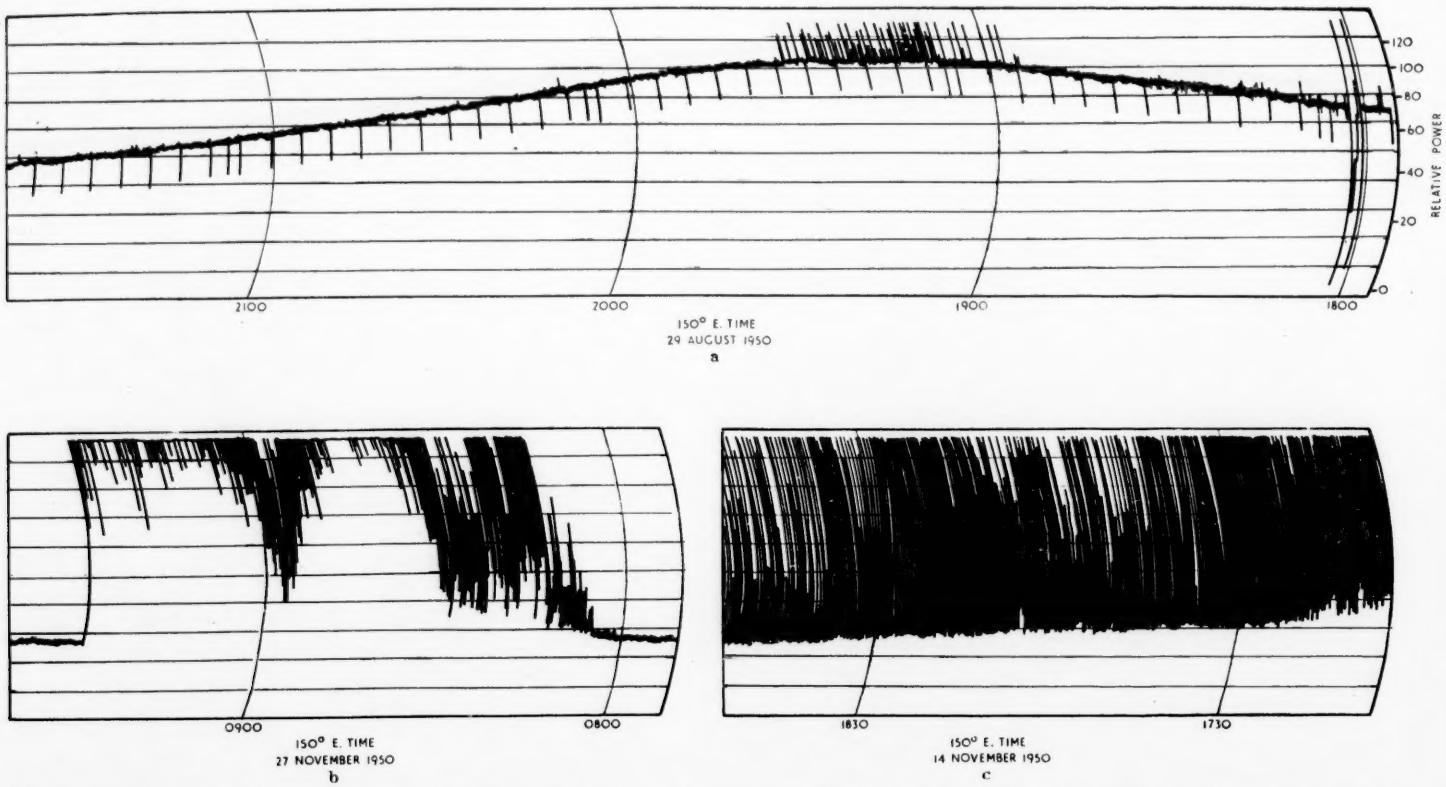


Fig. 1. Sample records. (a) A normal record. The broad maximum is recorded as the plane of the Galaxy (the Milky Way) passes through the aerial beam. Slight interference is visible from a radio station at about 1815 hours and from atmospherics at about 1920 hours. Calibration signals have been recorded at 1800 hours.
(b) and (c) Exceptional records. (b) Spoilt by strong interference from radio station. (c) Showing the end of a period of severe atmospherics. After about 1830 hours the base level can be detected and measured between the bursts of atmospherics.
It is possible to recognize and reject records showing interference which effaces the cosmic noise base level.

VOL. 4 1953

It was found that records taken at night (when the D, E, F_1 regions are absent) could be repeated to within the reading accuracy, from night to night, if the critical frequency of the F_2 region, f_0F_2, was in each case less than about 5 Mc/s. Received powers at other times were never higher. On a few occasions somewhat lower readings were obtained at night even when f_0F_2 was less than 5 Mc/s. As these occasions were infrequent they were considered exceptional and it was assumed that the records taken at night with f_0F_2 less than 5 Mc/s, and which gave values which could be repeated over several nights, could be used to establish the values of P_0 for all sidereal times and for all aerial directions considered. For all days in the period chosen for analysis of the records, the intensity was scaled from the records for each hour of the day (150°E. time), the corresponding sidereal time was calculated, the value of P_0 read from the standard curves, and the absorption deduced.

3. Observations

A continuous series of observations was available from June 1950 to June 1951. All the records were used for the purpose of establishing the values of P_0, special weight being given to some more accurate records taken at intervals throughout the year. To obtain some idea of the variations of absorption for different seasons without analyzing all the records, certain representative months were taken and their records analyzed for absorption effects. The periods chosen are set out in Table 1.

Table 1

Month	*Period of observations*
July	13 July–3 August
September	8 September–29 September
November–December	13 November–29 December

For the purpose of analysis, data from the nearest ionospheric recording station at Canberra (latitude 35°S., longitude 149°E.), about 250 km south-west of Hornsby, were used. These data are published in *Ionospheric Predictions—Series D* prepared by the Ionospheric Prediction Service of the Commonwealth Observatory.

The results of the observations of absorption are shown in Fig. 2, where hourly means are plotted. As expected, the absorption varies with the time of day and also with season. (For several reasons the observations for July and September are not as accurate as those for November–December and most emphasis will be placed on the result for the latter period.)

The values of absorption shown in this figure are the sum totals of individual contributions, some of which may be negligible, from the ionospheric regions

D, E, Es, F_1, and F_2 (including its upper part, normally inaccessible), and perhaps a region beyond the F_2 region. However, as will be shown later, the major contributions are those from F_2 and D.

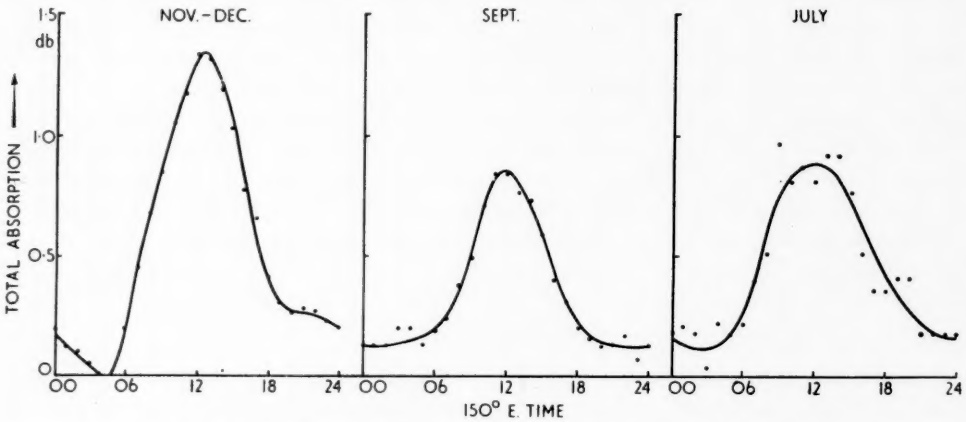


Fig. 2. Monthly mean hourly values of total observed absorption.

4. F_2 Region Absorption

4.1. *Evidence for F_2 region absorption, and its characteristics*

In the early stages of the analysis it was apparent that there was often appreciable absorption at night, and that at all times the absorption tended to increase with increasing values of f_0F_2. Fig. 3 shows a scatter diagram of absorption against f_0F_2 for all observations during the months of November–December. There is evidently some correlation between absorption and f_0F_2 and this was found even

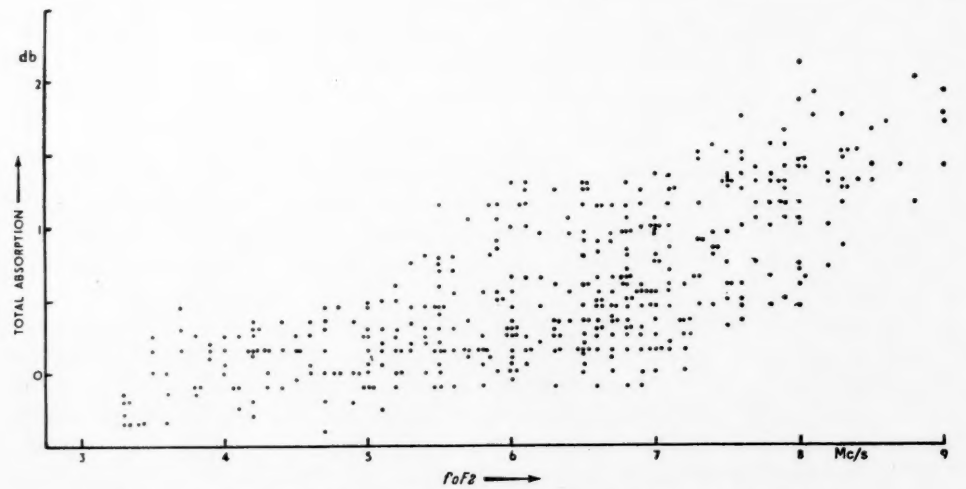


Fig. 3. Observed total absorption plotted against f_0F_2 (November–December).

VOL. 4 1953

when, as in Fig. 4, points for a single time of day were considered. This suggests that some, at least, of the absorption occurred in the F_2 region.

Such absorption seemed to depend on f_0F_2 alone, for, somewhat unexpectedly, no dependence of absorption on either of the heights recorded for the F_2 region

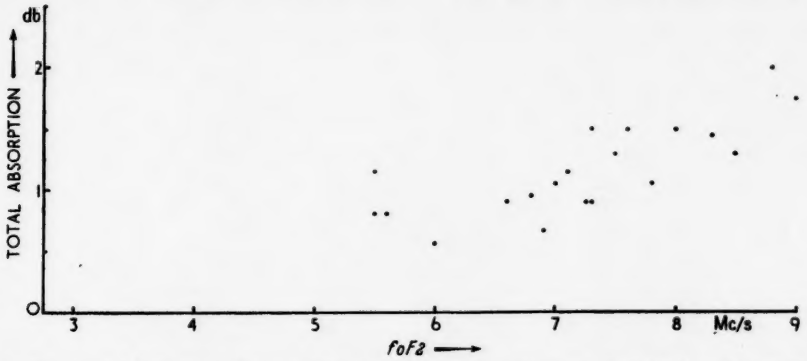


Fig. 4. Observed total absorption at 10 hours (November–December) plotted against f_0F_2.

(hpF_2 or $h'F_2$), or on their difference (hpF_2 *minus* $h'F_2$, which should be proportional to the scale height at the level of interest) could be detected. For example, Fig. 5 shows a plot of absorption against hpF_2 for a fixed solar time and

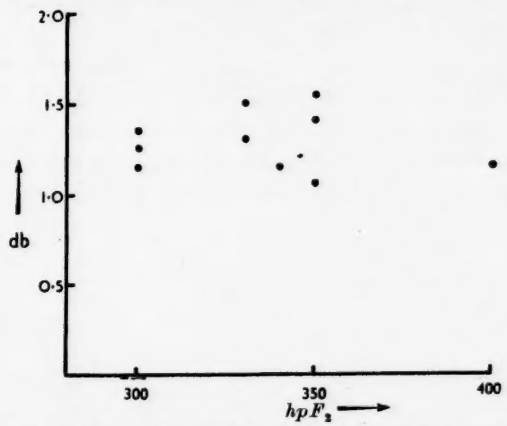


Fig. 5. Observed total absorption at 12 hours (for f_0F_2 between 7 Mc/s and 8 Mc/s) plotted against hpF_2 (November–December).

a small range of f_0F_2. Although only a small number of observations was used in Fig. 5, the fact that absorption is apparently independent of the height of the F_2 region has been confirmed by similar plots using night-time observations and also by comparing differences from the monthly means.

In order to find the magnitude of the F_2 region component of the total absorption, it is necessary to eliminate from the total absorption the possible contributions due to other regions. That such contributions exist is evident from the large scatter in Fig. 3 and also in Fig. 6, where the average values of absorption

for 1 Mc/s intervals of f_0F_2 are plotted against f_0F_2 for 06 hours, 08 hours and 11 hours solar time (November–December). It will be noted that the points for

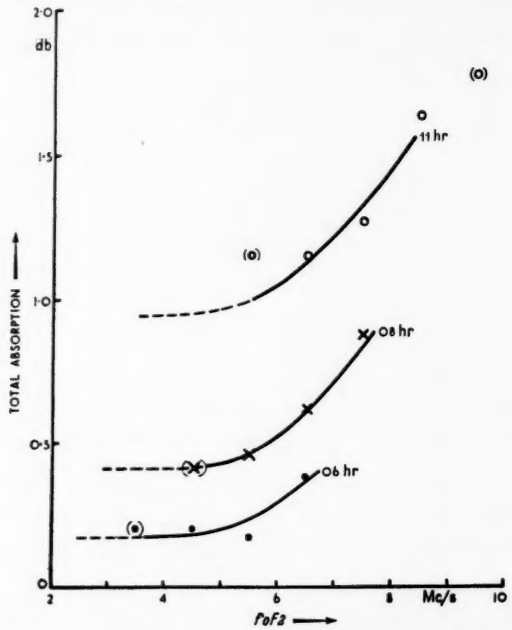


Fig. 6. Observed total absorption, averaged for intervals of 1 Mc/s of f_0F_2, for 06 hours, 08 hours and 11 hours plotted against f_0F_2 (November–December).

the three different times of day may be fitted by curves of the same shape but displaced progressively upwards. This implies that for each hour there is a component of absorption which is independent of f_0F_2 but depends on solar time.

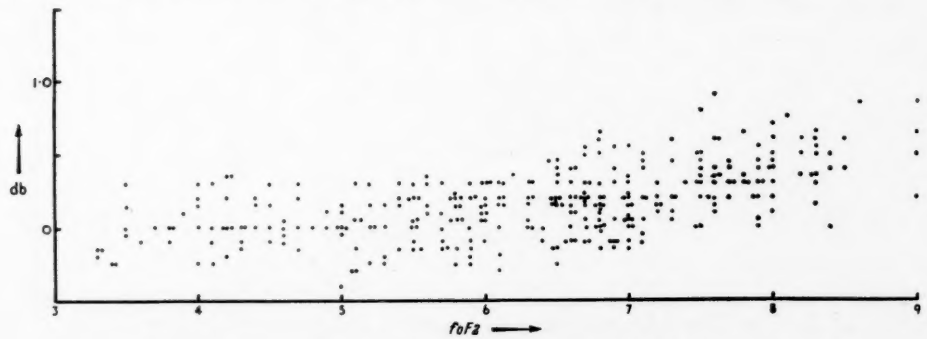


Fig. 7. F_2 region component of absorption plotted against f_0F_2 (November–December).

It is seen that the absorption becomes practically independent of f_0F_2 for low values of f_0F_2, and to find the absorption due to other regions it is necessary to determine the value of absorption that remains for low f_0F_2. In the early morning and late afternoon hours when fairly low values of f_0F_2 sometimes occur, the

VOL. 4 1953

extrapolation to no F_2 absorption involves little inaccuracy. However, during the middle of the day, for example at 11 hours (Fig. 6), f_0F_2 is seldom low and the slope of the curve of absorption against f_0F_2 is still appreciable for the lowest f_0F_2 recorded. In these cases the assumption was made that the shape of the absorption against f_0F_2 curve was constant throughout the day. In this way the

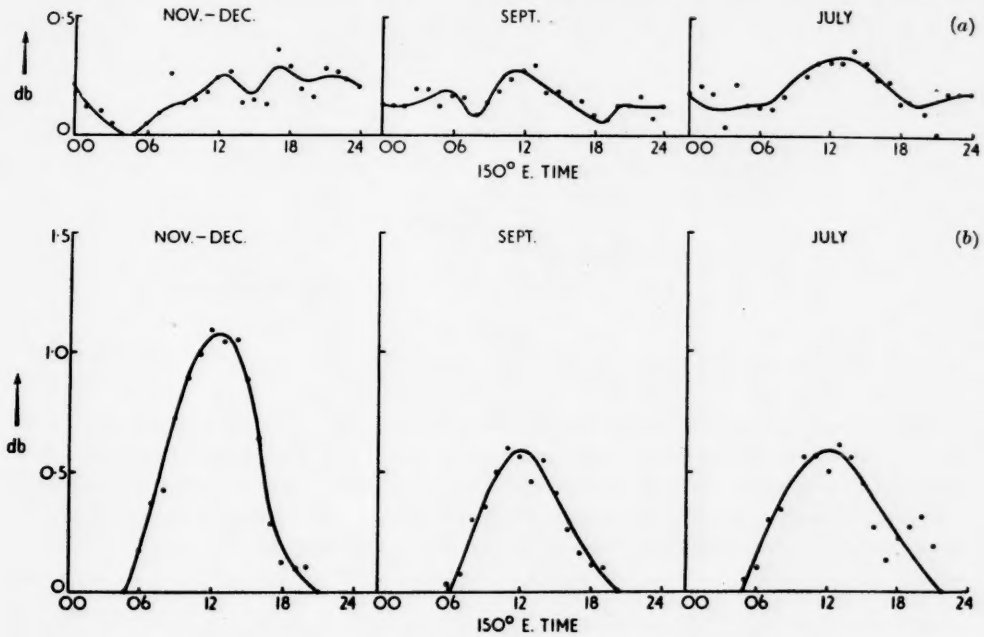


Fig. 8. Monthly mean diurnal variation of the two components of total absorption (*a*) F_2 (*b*) other than F_2 (shown to be largely D).

monthly mean residual absorption for each solar hour was estimated, and for each hour this amount was subtracted from all the values of total absorption observed at that hour. The remaining absorption, which is the component that may depend on f_0F_2, has been plotted against f_0F_2 in Fig. 7. Comparison with Fig. 3 shows that the scatter has been much reduced and the increase in absorption for the higher values of f_0F_2 is clearly defined.

The separated components of absorption for the three seasons are shown in Fig. 8. Fig. 8 (*a*) gives the diurnal variation of absorption due to the F_2 region, and 8 (*b*) the corresponding variations of absorption due to other regions.

In the course of analysis it was noticed that sometimes at night the absorption appeared to be unusually high for a particular f_0F_2, although no appreciable absorption would be expected from the lower layers of the ionosphere. As a check, the total night-time absorption and the F_2 component of the daytime absorption were each plotted against f_0F_2. In Fig. 9, the circles give the average total absorption for intervals of 0·5 Mc/s in f_0F_2 for observations made between 18 hours and 06 hours and the crosses the averages for the F_2 component for

observations between 06 hours and 18 hours (November–December data). It will be seen from Fig. 9 that for the higher values of f_0F_2 the absorption at night is, at least on the average, higher than during the day for the same value of f_0F_2. This phenomenon has been represented in another way in Fig. 10. In this figure,

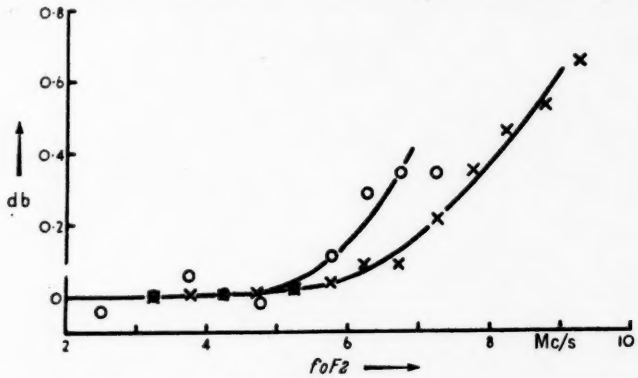


Fig. 9. F_2 absorption plotted against f_0F_2. Circles correspond to averages for observations during the night, crosses for observations during the day.

a comparison is made between the F_2 absorption observed and the absorption to be expected assuming that the average relationship between absorption and f_0F_2 for daylight hours were to hold at all times. It is seen that the discrepancy commences soon after sunset (about 19 hours) and is maximum at about 23 hours. It seems likely that the larger values of absorption at night are due to an additional source of ionization rather than a redistribution of the normal ionization, and this

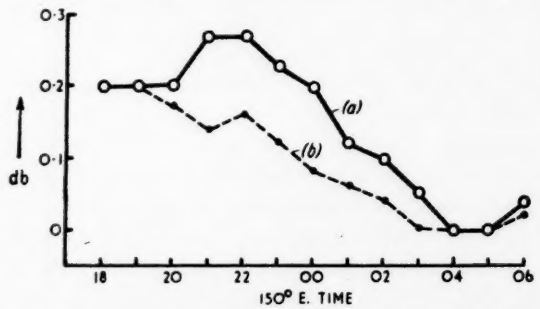


Fig. 10. F_2 absorption plotted against local time (November–December) showing excess absorption during the night. (*a*) observed values, (*b*) calculated values assuming the variation of F_2 absorption with f_0F_2 is the same during the night as during the day.

is confirmed by the fact that on several occasions there has been considerable absorption at night-time even when the corresponding values of f_0F_2 were very small.

A possible explanation of this extra absorption may be found in the ionospheric irregularities which cause fluctuations in the intensities of the discrete sources of cosmic noise and which are probably connected with the occurrence of "spread-F" echoes in normal ionospheric recordings (RYLE and HEWISH, 1950). In Fig. 11

VOL. 4 1953

are plotted the variations in the excess absorption for November and December 1950, taken from Fig. 10, as a function of time of day, and also the ratio, expressed as a percentage, of the number of occasions on which "spread-F" echoes were recorded to the total number of usable observations from the ionospheric recorder

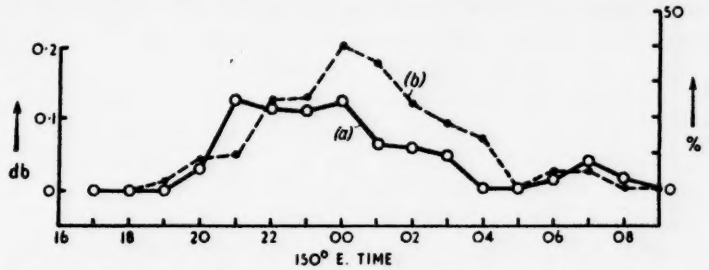


Fig. 11. (*a*) Excess night-time absorption, taken from Fig. 10. (*b*) Percentage occurrence of "spread-F" echoes.

at Canberra. It is seen that there is considerable similarity between the two curves and this suggests that scattering by irregularities of ionization in the upper F_2 region may be the cause of the excess absorption observed at night.

4.2. *Discussion—Variation of absorption with* f_0F_2

The absorption to be expected from a simple Chapman region for different values of the ratio of the operating frequency to the critical frequency of the region has

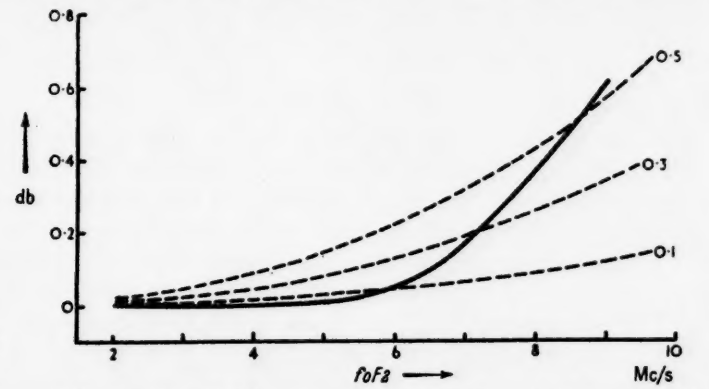


Fig. 12. Observed F_2 absorption (full line) compared with the theoretical absorption (dashed lines) for a Chapman region with several values of the parameter ($4{\cdot}34\ H\nu_m/c$).

been considered by APPLETON (1937) and JAEGER (1947), and typical curves are shown in Fig. 12, together with the observed variation obtained from Fig. 7. The theoretical curves are for three values of the parameter ($4{\cdot}34\ H\nu_m/c$) where H is the scale height of the region assumed constant throughout the region, ν_m is the collision frequency at the height of maximum ionization and c is the velocity of light. It will be seen that the curve of observed absorption has a different shape from the theoretical curve, showing a more rapid increase with critical frequency than the latter.

This result, obtained previously by Shain (1951), is not unexpected since previous ionospheric investigations have shown that the F_2 region shows little resemblance to a Chapman region, with large diurnal and seasonal anomalies in both the critical frequency and the equivalent height of the region. Modifications of the simple Chapman theory which must be considered include variation with height of the recombination coefficient and scale height and also the effect of large scale movements, for example those due to tidal forces, in the F_2 region. Nicolet's work (1951) on absorption under these more complicated conditions, although originally intended for application to the D region, must be extended to the case of the F_2 region. It seems doubtful, however, whether this will give a sufficiently rapid increase in absorption with critical frequency.

It is possible that scattering by irregularities may increase the absorption at all times but there is, at present, no evidence that these irregularities are important during daytime.

The incoming radiation will suffer attenuation due to refraction effects as f_0F_2 approaches the operating frequency. Although such attenuation, calculated by the methods of Bremmer (1949) should be small, it is possible, as was suggested previously (Shain, 1951), that the unexpectedly high F_2 region attenuation may be associated with the unexpectedly high refraction observed in the measurements of the positions of extra-terrestrial sources of radio waves (Payne-Scott and McCready, 1948; Kerr and Shain, 1951). Mitra (1952) has suggested that the large refraction effect may be due to an asymmetry in the ionization distribution in the F region caused by a variation in the recombination coefficient with height, giving a region of which the upper part is many times thicker than the lower part. The refraction attenuation due to such a region has not been calculated.

VOL. 4 1953

Lack of correlation of absorption with F_2 region heights

For a given value of the critical frequency of a simple Chapman region, the absorption varies as Hv_m. Although it has been shown that the Chapman theory is not valid in some respects for the F_2 region, it is probable that the proportionality of absorption to Hv_m should still be roughly correct. The collision frequency decreases with increasing height and it was expected that as the height of the F_2 region varied, the absorption would vary in the opposite sense. As this variation was not observed, it is probable that the product Hv_m is constant with height, at least to a first approximation, and hence that the scale height (and therefore temperature) increases with height over the range of heights concerned. Such a variation of temperature in the F_2 region is consistent with present data on this subject (Gerson, 1951).

5. D Region Absorption

The component of absorption shown in Fig. 8 (b) may include separate components from the D, normal E, sporadic E and F_1 regions, although contributions from the regions other than the D region would be expected to be small. As before, scatter diagrams have been drawn of absorption against f_0E, fEs, and f_0F_1 for a particular solar hour and for a small range of f_0F_2. These are not reproduced but it was found that there was no correlation between absorption and these three

ionospheric parameters, although in the case of the E and F_1 regions, the values of f_0E and f_0F_1 vary little from day to day for a given solar hour, so that small contributions from these regions would not be detected. However, it may be concluded that at least the major part of the absorption not due to the F_2 region is due to the D region. This is in agreement with the conclusions of APPLETON and PIGGOTT (1948) from pulse reflection experiments.

For comparison with the observations made at Slough and elsewhere it is convenient to calculate the absorption index A. This is defined as the most probable value of the non-deviative absorption in decibels at vertical incidence for an effective frequency (that is $f \pm f_L$, where f is the frequency of observation and f_L is the gyro-frequency corresponding to the component of the earth's magnetic field along the ray path) of 1 Mc/s. The values obtained at noon are compared in Table 2 with the values recorded at Slough (latitude 52°N., longitude 1°W.) and at Port Stanley, Falkland Islands (latitude 52°S., longitude 58°W.).* For Slough two values are given for the summer and winter periods to permit comparison with the corresponding intermediate southern hemisphere seasons. There were no absorption data for the Falkland Islands for November and December 1950 and January 1951.

Table 2. Noon values of A(db)

Station	*Summer*	*Equinox*	*Winter*
Slough	440 May–June 1950 510 May–June 1951	480 September 1950	390 January 1950 490 January 1951
Falkland Islands	360 October 1950 380 February 1951	300 September 1950	280 July 1950
Hornsby	390 November–December 1950	200 September 1950	200 July 1950

It is seen that although the Slough values are considerably higher in all seasons, the Hornsby cosmic noise results for summer (which are the most reliable) agree well with those recorded at the Falkland Islands, while the trend with season is also similar. The cause of the difference between Slough and the other stations is unknown—all three stations have roughly the same magnetic latitude.

A full consideration of the seasonal variation would require analysis of many months' data, but it would appear that the cosmic noise technique could be used to provide suitable data from many stations in different parts of the world.

It has been found from earlier pulse measurements that the curve of D region absorption against time of day is usually roughly symmetrical about a time slightly later than local noon. BRACEWELL and BAIN (1952) quote a private communication from PIGGOTT, giving a value, deduced from the Slough absorption data, of 33 minutes for the lag of the time of symmetry after local noon. Also

* The authors are grateful to the Director of Radio Research, D.S.I.R., for permission to use the data from the Monthly Bulletins of Ionospheric Characteristics issued by the D.S.I.R. Radio Research Station, Slough.

BRACEWELL (1952) has observed a diurnal asymmetry in the characteristics of the propagation of 16 kc/s radiation over distances of a few hundred kilometres which he attributes to the variations in a layer which probably causes the absorption of higher frequency radio waves. He found a time lag of 25 minutes.

Examination of Fig. 8 (*b*) shows that with the present observations the smoothed maximum value of the *D* region component of the absorption occurs after noon, the effect being particularly noticeable in the summer. The effect has been studied quantitatively as follows. The times were found, one before and one after noon, at which the absorption passed through a particular value. The time half way between these two times was calculated and the differences obtained between this time and the time of transit of the sun. This procedure was repeated for several values of absorption and the mean taken. It was found that for each month the time lags were of the same order of magnitude, the most reliable value, for November–December, being 27 minutes. This time lag is in good agreement with values found in the other investigations.

It is known from previous absorption observations that the diurnal variations of *D*-region absorption is approximately of the form $\cos^n \chi$, where χ is the solar zenith angle, and it is of interest to find the value of n which would best fit the present observations. This has been done for the three months separately, in each case allowing for the lag of the time of symmetry with respect to local noon. The values obtained for n are shown in Table 3, together with values deduced from the published Slough data for January 1949 to June 1949, and also the values deduced by TAYLOR (1948) from observations of continuous waves at practically vertical incidence. The agreement between the present results and the other observations is satisfactory.

VOL. 4 1953

Table 3. *Values of the exponent of cos* χ

Author	*Frequency (Mc/s)*	*n*
Taylor	2·061	1·00 summer 0·87 equinox 0·73 winter
(Slough)	1·6–5·5	0·9 summer 0·8 equinox 0·6 winter
Present observations	18·3	1·1 summer 0·9 equinox 0·5 winter

There is still some doubt as to the properties of the *D* region and further experimental work, which must necessarily include observations of absorption, is required, preferably on a world-wide basis, for testing theoretical conclusions.

For example, it is known (SMITH-ROSE, 1951) that although the diurnal variation of absorption is of the form $\cos^n \chi$, the seasonal variation of noon absorption will not fit such a simple relationship. Further light may be thrown on this matter if absorption observations could be taken on the same days at several stations in different latitudes for which the noon value of $\cos \chi$ may vary over a wide range. The present simple method of measuring absorption may make such observations practicable.

6. CONCLUSION

A new technique has been described which enables easy and accurate measurements to be made of the total attenuation suffered by a radio wave passing completely through the ionosphere, and it has been shown that two main contributions, from the D and F_2 regions, can be recognized and separated.

At present the number of stations making measurements of ionospheric absorption is small compared with the number making routine recordings of critical frequencies and virtual heights. The simplicity of the present technique, which also permits continuous observations of absorption and hence measurements of sudden ionospheric disturbances and other cases of abnormal absorption, should enable many observing stations to be operated throughout the world using similar equipment and this would allow ready comparison between the observations from different stations.

One of the most important applications of the technique would be the operation of suitable equipment for measuring the total absorption of cosmic noise alongside pulse equipment for measuring absorption by reflection. The comparison of the two kinds of records should give valuable information concerning the absorption of radio waves in the F_2 region, particularly in the important region above the level of maximum ionization which is inaccessible to the normal reflection technique.

The present study confirms the results of previously published observations but suggests that further theoretical work is required to interpret the discrepancies which exist between observational results and present theory, with particular regard to the F region.

Acknowledgements—The authors are indebted to Mr. C. A. HIGGINS for much assistance in the taking of the observations discussed in this paper and in some of the analysis. During the investigation one of the authors (A.P.M.) was the holder of a Fellowship awarded by the Australian Government under the Colombo plan.

REFERENCES

APPLETON, E. V.	1937	Proc. Roy. Soc. **A162,** 451–479
APPLETON, E. V., and PIGGOTT, W. R.	1948	Proc. Gen. Assy. U.R.S.I. Stockholm, 320–322
BRACEWELL, R. N.	1952	Proc. I.E.E. **99,** Part 4, 217–228
BRACEWELL, R. N., and BAIN, W. C.	1952	J. Atmosph. Terr. Phys. **2,** 216–225
BREMMER, H.	1949	"Terrestrial Radio Waves," Elsevier Publishing Co. Inc., New York
GERSON, N. C.	1951	Reports on Progress in Physics **14,** 316–365

JAEGER, J. C.	1947	Proc. Phys. Soc. **59,** 87–96
JANSKY, K. G.	1937	Proc. I.R.E. **25,** 1517–1530
KERR, F. J., and SHAIN, C. A.	1951	Proc. I.R.E. **39,** 230–242
LAFFINEUR, M., and KOURGANOFF, V.	1950	Proc. Gen. Assy. U.R.S.I. Zurich, Part 2, 261–265
MITRA, A. P.	1952	Indian J. Phys. **26,** 495–511
NICOLET, M.	1951	J. Atmosph. Terr. Phys. **1,** 141–146
PAYNE-SCOTT, R. and MCCREADY, L. L.	1948	Terr. Mag. **53,** 429–432
RYLE, M., and HEWISH, A.	1950	Mon. Not. Roy. Astr. Soc. **110,** 381–394
SHAIN, C. A.	1951	Aust. J. Sci. Res. **A3,** 258–267
SHAIN, C. A., and HIGGINS, C. A.	1954	"Observations of the general background and discrete sources of 18·3 Mc/s cosmic noise," (in preparation)
SMITH-ROSE, R. L.	1951	D.S.I.R. "Radio Research 1950," London H.M.S.O., 7–47
TAYLOR, E. W.	1948	Bur. Standards J. Res. Wash. **41,** 575–579

VOL.
4
1953

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 219 to 225. Pergamon Press Ltd., London

The strength of meteoric echoes from dense columns*

LAURENCE A. MANNING

Electronics Research Laboratory of the Department of Electrical Engineering
Stanford University, Stanford, California

(*Received* 7 *October* 1953)

ABSTRACT

Calculations of the strength of the echoes from dense meteoric ionization columns have been made in the past by assuming that the trails behave like perfectly reflecting cylinders of that radius for which the refractive index is zero. It is shown that this procedure neglects the defocusing effect of refraction at greater radii. A general formula is obtained for the effect of refraction on echo strength. The maximum power returned from a large dense Gaussian distribution of ionization is found to be only 70 per cent of that assumed previously.

INTRODUCTION

Echoes from meteoric ionization columns have long been known to differ greatly from one another in amplitude and duration. Relatively recently there have appeared theories designed to explain in detail the causes of these differences in terms of a plausible ionization distribution model (KAISER, 1952; ESHLEMAN, 1952; GREENHOW, 1952). Because of the excellent confirmation with experiment, there can now be little doubt that a meteor produces a concentrated cylindrical distribution of ionization, confined to an initial radius of only a few centimetres. As time passes, the density of the ionization along the trail axis decreases and the radial size of the distribution is increased by diffusion. Given such a model, it is but an analytical problem to compute the strength of the echoes that would be sent back to a sounding station.

Calculation of the intensity of meteoric echoes has been carried out so far by finding approximate formulae that are good in restricted ranges of ionization density and size. For weakly ionized trails, exact solutions have been obtained. For large, highly ionized trails, it has been assumed that the reflection coefficient is the same as would result from a metallic cylinder of size equal to the radius of zero refractive index. The solutions for these extreme cases differ in a number of ways. For the weakly ionized trails, the reflected signal strength decays essentially exponentially in time; the rate of decay is determined by the diffusion coefficient, but not by the initial density. For the highly ionized trails, the critical density radius rises to a maximum value at $1/e$ of the total duration; the duration increases with rising initial density. By dealing only with the maximum amplitudes in the two cases, curves of maximum reflected signal can be plotted as a function of the initial line density of charge. In the transition region, bridging formulae can be used.

The objection may be made that use of the critical-density-radius for highly ionized trails neglects the refractive effects produced by the ionization distributed at greater radii. Unfortunately, the exact calculation of the high density case is very difficult. However, the use of geometrical optics and ray tracing methods is

* Jointly supported by the U.S. Army Signal Corps, the U.S. Air Force, and the U.S. Navy (Office of Naval Research)

applicable to sufficiently large columns. In this paper their application will be made to the calculation of the reflection coefficient for the dense case. The results will be shown to differ from those computed heretofore.

GEOMETRY OF THE REFLECTION PROCESS

In developing the general theory, it will be assumed only that there is a cylindrically symmetrical distribution of refractive index about the trail axis.

As in Fig. 1 suppose that refractive index is a function of r alone. As r becomes very large, the refractive index will approach the free space value of unity. An incident ray will be directed towards the column in such a way as to miss the axis

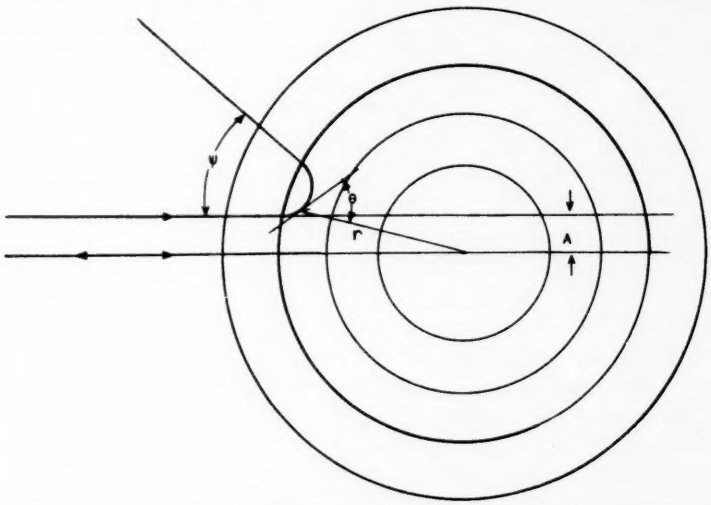


Fig. 1. Ray paths in a large over-dense cylindrical ionization column.

by a distance A. At each point of the ray path, the radius vector from the cylinder axis of length r makes an angle θ with the direction of the incident ray. In accordance with BOUGER'S rule

$$\mu(r) \sin \theta = A \tag{1}$$

where $\mu(r)$ is the refractive index. For a ray of position fixed by A, the direction of travel and hence the path can be found by integrating equation (1). The quantity we wish to compute, however, is the reflected signal strength. It may conveniently be specified in terms of the radius of the equivalent metallic cylinder that would give the same echo as is observed. Suppose a ray to be reflected as shown in Fig. 1. The energy contained between this ray and one directed at the axis is proportional to their separation A. The reflected energy is fanned out over an angle ψ. As A is reduced, ψ simultaneously approaches zero. The limiting ratio $d\psi/dA$ is a measure of the defocusing or spreading of the reflected wave. In the case of a metallic cylinder the radius ρ is related to the spreading by the expression

$$\rho = 2/\left.\frac{dA}{d\psi}\right|_{A=0} \tag{2}$$

VOL. 4 1953

In the case of a refracting ionized column, the radius of the equivalent metallic cylinder may be found from (2) by computing the deviation as A approaches zero, using (1).

Ray Path Calculation

The deviation in an ionized column of an incident ray of position A will now be found. Referring to Fig. 2, note that $\tan\theta = r\mathrm{d}\psi/\mathrm{d}r$. As a consequence,

$$\psi = 2\int_R^\infty \frac{\tan\theta\,\mathrm{d}r}{r}$$

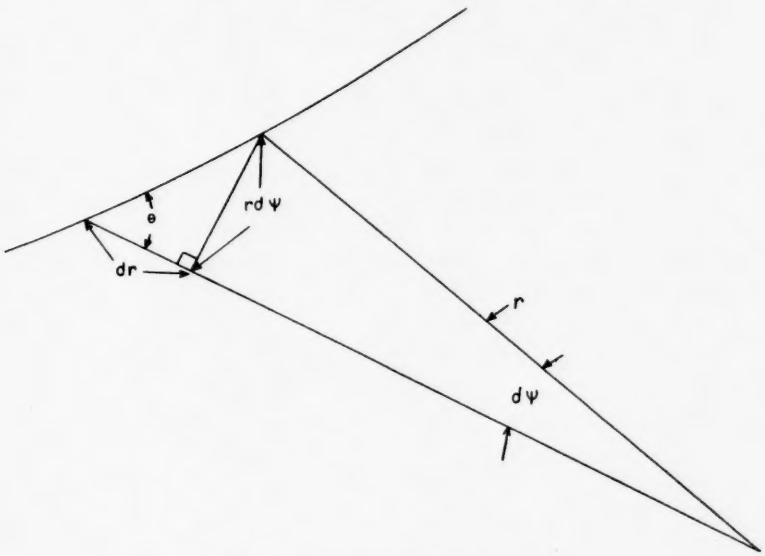


Fig. 2. Geometry of the wave path.

where R is the closest approach of the ray to the cylinder axis. From (1), θ may be eliminated in favour of r and A. If in addition we introduce the normalization $r = Rx$,

$$\psi = 2\int_1^\infty \nu\frac{\mathrm{d}x}{x} \tag{3}$$

where

$$\nu = 1/\sqrt{\frac{\mu^2(Rx)R^2x^2}{A^2} - 1} \tag{4}$$

Upon inspecting (3), it will be seen that only ν is a function of A on the right-hand side. To find $\mathrm{d}\psi/\mathrm{d}A$, and so the spreading, we must evaluate $\mathrm{d}\nu/\mathrm{d}A$. We find

$$\frac{\mathrm{d}\nu}{\mathrm{d}A} = \frac{\mu^2(Rx)R^2x^2 - A\mu^2(Rx)\,x^2R\dfrac{\mathrm{d}R}{\mathrm{d}A} - A\mu'(Rx)x^3R^2\dfrac{\mathrm{d}R}{\mathrm{d}A}}{(\mu^2R^2x^2 - A^2)^{3/2}} \tag{5}$$

Since we wish to evaluate $d\psi/dt$ for $A = 0$, we must evaluate (5) for $A = 0$. To do this, we must find dR/dA for $A = 0$. From (1), note that $r = R$ when $\sin \theta = 1$. Hence

$$R\mu(R) = A$$

and

$$\frac{dR}{dA} = 1/(R\mu'(R) + \mu) \tag{6}$$

As A approaches zero, $\mu(R)$ approaches zero. Assuming μ to be caused by ionization.

$$\mu = \sqrt{1 - \frac{KN(R)}{f^2}}$$

and

$$\frac{d\mu}{dR} = \tfrac{1}{2}\frac{(-KN'(R)/f^2}{\mu(R)}$$

so that if $N(r)$ has a finite derivative, $d\mu/dr$ becomes infinite as A, hence $\mu(R)$ approaches zero. Referring again to (6) as A goes to zero, so then does dR/dA. Considering (5) again as A and dR/dA approach zero, note that the two final terms in the numerator vanish. The first numerator term then cancels with the reduced denominator, and

$$\frac{dv}{dA} = \frac{1}{Rx\mu(Rx)} \tag{7}$$

for $A = 0$.

Referring to (3), we find

$$\frac{d\psi}{dA} = 2\int_1^\infty \frac{dv}{dA}\frac{dx}{x}$$

and for $A = 0$,

$$\left.\frac{d\psi}{dA}\right|_{A=0} = 2\int_1^\infty \frac{dx}{Rx^2\mu(Rx)} \tag{8}$$

Behaviour of the Equivalent Column Radius

Equation (8), in combination with (2), yields the general expression for the radius of the metallic reflector returning the same back-scattered signal as a large overdense cylindrical column of ionization. Recalling that R is the critical density radius for back reflection, and ρ is the radius of the equivalent metallic cylinder,

$$\frac{R}{\rho} = \int_0^1 \frac{dy}{\mu(R/y)} \tag{9}$$

where $y = 1/x$.

Since the refractive index will always be less than or equal to one, it will be seen that the equivalent cylinder size must always be less than or equal to the critical density radius.

VOL. 4 1953

The example of greatest interest is the diffusion formed trail. If it is assumed that Q electrons are initially deposited along unit length of the trail, and that the diffusion coefficient is D, the electron density N will be

$$N = \frac{Q}{4\pi Dt} \exp(-r^2/4Dt) \tag{10}$$

at time t. The refractive index is given by $\mu^2 = 1 - KN/f^2$. The duration of the

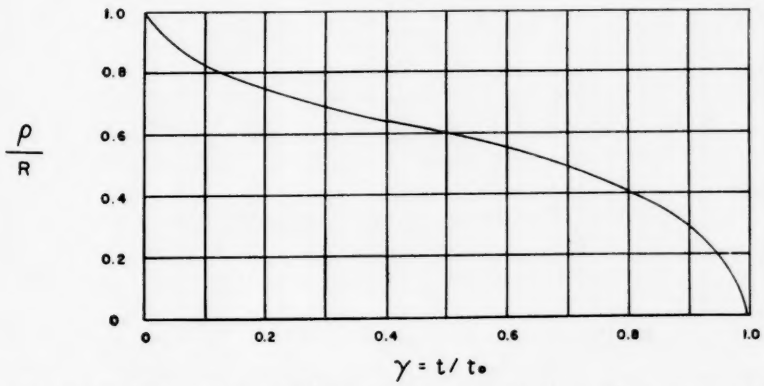


Fig. 3. Ratio of the effective radius of a dense diffusion formed ionization trail to the critical density radius versus time.

echo is found by solving (10) for r, and letting the critical-density-radius vanish. For the critical-density-radius $r = R$, $\mu = 0$, and $N = f^2/K$. From (10)

$$r^2 = 4Dt \ln Q/4\pi DtN$$

so that if t_0 is the duration, $4\pi Dt_0 f^2 = KQ$. From the preceding, and (10), the refractive index may be written

$$\mu^2 = 1 - \frac{1}{\gamma} \exp(-r^2/4Dt_0\gamma)$$

When $r = R$, $\mu = 0$ and $\gamma = \exp(R^2/4Dt_0\gamma)$. Since $r = R/y$,

$$\mu^2 = 1 - \frac{1}{\gamma} \exp \frac{-\ln(1/\gamma)}{y^2} \tag{11}$$

and

$$\frac{R}{\rho} = \int_0^1 \frac{dy}{\sqrt{1 - \frac{1}{\gamma} \exp \frac{-\ln(1/\gamma)}{y^2}}} \tag{12}$$

Fig. 3 shows a plot of (12), the integration having been performed with a planimeter. Notice that early in the lifetime of the echo the effective radius differs but little from the critical-density-radius. However, as the trail expands, a larger amount of ionization exists outside the critical radius, and the refractive effects are augmented. Near the end of the echo, the ratio of the effective to critical

radii approaches zero, and the inadequacy of the metallic-cylinder critical-density-radius method is most pronounced. In general, as can be seen from (9), the more diffuse the ionization distribution, the more important will be the refractive effects.

Fig. 3 expresses the equivalent radius of the diffusion formed trail in terms of the critical-density-radius. It may be shown that the critical-density-radius R, in terms of the maximum value R_m, is given by the expression

$$\frac{R}{R_m} = \sqrt{e\,\gamma \ln (1/\gamma)} \tag{13}$$

where $\gamma = t/t_0$.

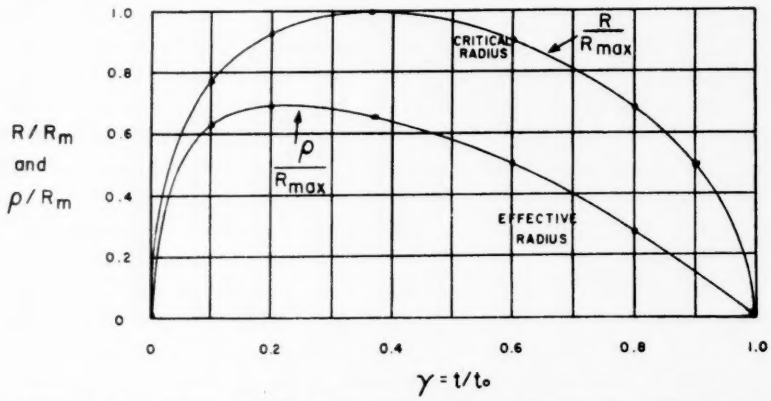


Fig. 4. Time behaviour of the critical density radius and the effective radius of a dense Gaussian ionization trail.

In Fig. 4 are compared the critical-density and equivalent radius time variations. It will be seen that for a Gaussian trail refractive effects reduce the maximum received echo power to 70 per cent of the value computed in published calculations of meteoric echo strength. In addition, the amplitude function is altered so that the maximum echo strength occurs nearer the start of the echo.

CONCLUSIONS

Calculations of the strength of echoes from dense meteoric ionization columns have been made in the past by assuming the trail to act like a metallic cylinder of the radius for which the refractive index becomes zero. It has been shown, however, that if refractive effects are considered, the maximum received power from a diffusion formed trail will be reduced to 70 per cent of that computed using the critical radius method. For more general continuous distributions of ionization, the critical-density-radius method will always give a high answer. The extent of the error which results from its use depends upon the amount of ionization exterior to the critical-density-radius. Although the present calculations appear rigorous for very large columns the assumptions of geometrical optics on which they are based are not upheld for small columns. In their treatment of the dense column case, KAISER and CLOSS effectively used the zero refractive index radius

VOL. 4 1953

only for large trails; for the small radii dense trails, they chose an effective radius where the dielectric constant was progressively more negative. In this way a smoother transition was obtained to the scattering formulae. The refractive effects considered here may be expected to be important in the upper density region for which KAISER and CLOSS' assumption approaches that of the zero refractive index radius.

REFERENCES

ESHLEMAN, V. R. 1952 "The Mechanism of Radio Reflections from Meteoric Ionization," Ph.D. Dissertation, Dept. of Electrical Engineering, Stanford University

GREENHOW, J. S. 1952 "Characteristics of Radio Echoes from Meteor Trails: III The behavior of the Electron Trails after Formation," Proc. Phys. Soc. B **65**, 169–181

KAISER, T. R., and CLOSS, R. L. 1952 "Theory of Radio Reflections from Meteor Trails: I," Phil. Mag. **43**, 1–32

DL.
4
953

Journal of Atmospheric and Terrestrial Physics. 1953, Vol. 4, pp. 226 to 239. Pergamon Press Ltd., London

The $H\alpha$ radiation from solar flares in relation to sudden enhancements of atmospherics on frequencies near 27 Kc/s

M. A. ELLISON

Royal Observatory, Edinburgh

(*Received* 15 *October* 1953)

ABSTRACT

The development curves of 10 solar flares observed in $H\alpha$ light are compared with the sudden enhancements of atmospherics (S.E.A.s) generated by them. In confirmation of earlier work (1950), an average time lag of 7 minutes is established between the maximum of the flare and the maximum of the resulting S.E.A. This is interpreted as an effect of "sluggishness" in the response of the *D*-layer (at a height of about 75 km) to the ionising radiation from the flare.

An analysis is made of the association of S.E.A.s with 413 flares of Classes 1, 2 and 3 during a period of 4 years. In the hours between 0900 and sunset 51 per cent of Class 1 flares, 88 per cent of Class 2 flares and all Class 3 flares had associated S.E.A.s. The percentage association is higher during the middle of the day and in the afternoon than in the early morning hours. It is also higher in winter than in summer.

It is emphasised that three or four 27 Kc/s recorders well distributed in longitude would give almost complete coverage for the recording of major flares; the method might profitably be used during the International Geophysical Year of 1957.

1. INTRODUCTION

In an earlier publication (ELLISON, 1950) some detailed comparisons were made between the $H\alpha$ development curves of 13 well observed solar flares and the sudden ionospheric disturbances (S.I.D.s), generated simultaneously by the flare ultraviolet light as it illuminated the lower ionosphere. The following types of S.I.D.s were considered in relation to the flares:

(1) Crochets in terrestrial magnetism (conductivity effect).

(2) Radio fadeouts 5–20 Mc/s (absorption effect).

(3) Sudden phase anomalies of very long radio waves at a frequency of 16 Kc/s (lowering of *D*-layer reflecting ceiling).

(4) Sudden enhancements of long-wave atmospherics recorded on a frequency of 27 Kc/s (improved reflectivity of *D* at oblique incidence).

(5) Sudden anomalies in the field strength of a distant 16 Kc/s transmitter (interference between ground-wave and sky-wave from *D*).

Each type of S.I.D. was found to have its own characteristic form of development curve, and all differed somewhat from that of the $H\alpha$ light, though it was evident that each was uniquely related to the flash of radiation which occurred near the time of commencement of the flare. In particular, it was discovered that the maximum of the extra-ionisation in *D* (as shown by effects 3, 4 and 5) was reached on the average about 6 minutes later than the time of peak intensity of the flare. Thus, it appeared that the electron density *N* in the *D* region lagged behind the progress of the visible radiation. During the decay phase this effect was even more pronounced, so much so that, in the case of the most intense flares, the *D*-layer did not recover completely for an hour or more after all visible radiation in $H\alpha$ had ceased. It is, of course, natural to suppose that the intensities of the visible and ultraviolet wavelengths in the flare radiation follow the same

VOL.
4
1953

course of rise and decay with time, even though only the former can at present be directly observed. At all events, this assumption will be made until there is evidence to the contrary.

These facts, being of great interest to ionospheric studies, made it very desirable to obtain more numerous observations and more precise information about the course of flare development, especially in the early stages which are the most difficult to observe. In the past three years many more flare curves have become available, and in the present paper some of these are considered in relation to the sudden enhancements of atmospherics recorded in Edinburgh on frequencies between 27 and 22 Kc/s (11,000–14,000 metres).

An ideal ionospheric experiment might be performed in the following way. We could arrange to illuminate a given layer with an *instantaneous flash* of ultraviolet light, whose peak intensity was, say, ten times the normal value. We would then record by radio methods the build-up and subsequent decay of the electron density. Although such an experiment is beyond our powers of achievement by artificial means, nature, through a remarkable mechanism occurring in the solar chromosphere, performs a similar operation for us on many occasions. As we shall see, the flare development curves tell us much about the rate at which the ultraviolet light is "switched on" and "switched off"; and the S.I.D.s indicate the course of the resulting ionisation in the D region, provided only that we can interpret the radio messages on different frequencies correctly.

OL.
4
953

2. Observational Methods

2.1. *Hα development curves*

These graphs (continuous lines of Fig. 4) show the *effective width* of the flare $H\alpha$ emission line plotted against time. Five of the curves were obtained by measurements made with the spectrohelioscope at the Royal Observatory, Edinburgh,

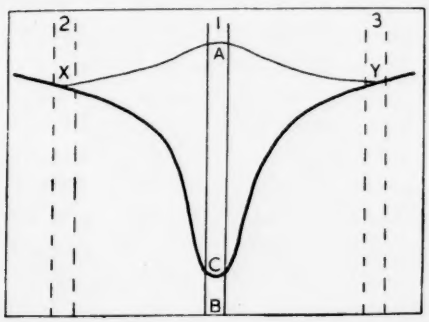


Fig. 1. Fraunhofer profile (XCY) and flare profile (XAY) in $H\alpha$ light.

and five with a similar instrument at the observatory of Ondřejov, Czechoslovakia. The latter are reproduced by kind permission of Drs. F. Link and D. Mášková, from their valuable catalogues (Link and Mášková, 1951; 1952) of solar flare observations. Precisely similar methods were used at both observatories. Briefly, in Fig. 1, XAY represents the $H\alpha$ emission profile of a moderately intense flare,

and XCY is the profile of the normal Fraunhofer $H\alpha$ line of the solar chromosphere upon which the emission is superimposed. Then XY, which may be called the *effective width* of the emission line, is measured by setting the second slit of the spectrohelioscope (here drawn to scale) successively in the positions 1, 2 and 3. The slit is set first in position 1, for darkest field (and usually brightest for the flare), then in positions 2 and 3, the points at which the emission loses contrast with its background. To avoid systematic errors the slit is moved in the outward direction, from 1 to 2 and from 1 to 3. The *effective width* is, therefore, the same thing as the *visibility range* of the emission line and is measured in angstroms. Such measures can be made at a rate of one or two per minute during the course of a flare. Alternatively, the central intensity, AB, of the line may be determined with a photometer (NEWTON, 1939; ELLISON, 1952) and may be expressed either in terms of the brightness of the continuous spectrum outside the wings of the Fraunhofer line, or in terms of the brightness, BC, of the normal chromosphere. Such a series of measures (corrected for instrumental scattered light) is given in Fig. 4 (No. 8, I.C.) alongside the line-width curve. Experience has shown that the line-width of $H\alpha$ increases and decreases rather more rapidly than the central intensity, although the times of beginning, maximum and end of the flare are in close agreement whichever method is employed to record the development.

Ideally, of course, one would like to be able to follow the *total energy output* of a flare in $H\alpha$ light, i.e. to obtain continuous measures of the areas $XAYC$ in equivalent angstroms between the two profiles. Such measures have been made on a number of flares (ELLISON, 1952), but since they involve precise photographic photometry and the application of troublesome corrections the method is not yet applicable to high-speed work. In practice, therefore, the determination of line-width remains the simplest and most reliable means of following the development of a flare.

VOL.
4
1953

2.2. *Sudden enhancements of atmospherics*

It is now generally accepted that the flash of ultraviolet light generated in a flare produces a layer of extra-ionisation in the D region. The absorption coefficient for short radio waves traversing this layer experiences a sudden increase and this gives rise to the fadeouts, either partial or complete, on long distance channels of communication, phenomena which were first observed by MÖGEL (1930) (see also APPLETON, 1938) and were later associated with flares by DELLINGER (1935). In addition, the coefficient of reflection for very long radio waves, incident obliquely upon the base of the D-layer, is raised, which in turn produces large enhancements of the signals received after successive reflections from a distant transmitter. R. BUREAU (1937), to whom we owe this discovery, has investigated these enhancements by continuously recording thunderstorm atmospherics at many different wavelengths. He finds that the sudden enhancements are characteristic of the spectral range 7 km to 16 km (BUREAU, 1947), being most pronounced at a wavelength of about 11 km (27 Kc/s). There is no question of an increase in the number of atmospherics at such times; the phenomenon is simply one of improved propagation at these wavelengths.

The frequency of occurrence of S.E.A.s shows a fairly close dependence upon

the general level of solar activity throughout the 11-year cycle. This has been demonstrated statistically by BUREAU (1941) in relation to both the relative sunspot number (R) and the areas of calcium *plages faculaires*; and by J. KLECZEK (1952) in relation to a flare activity index (Q) depending upon the importance and duration of flares observed each day. At the Royal Observatory, Edinburgh, S.E.A.s have been recorded continuously since June 1949. The object here has been to make comparisons between individual flares and their S.E.A.s, as well as to investigate the suitability of this means for detecting flares and estimating their importance when direct observation is impossible. These results, obtained during four years (1949–53) of declining solar activity, form the basis of the present paper.

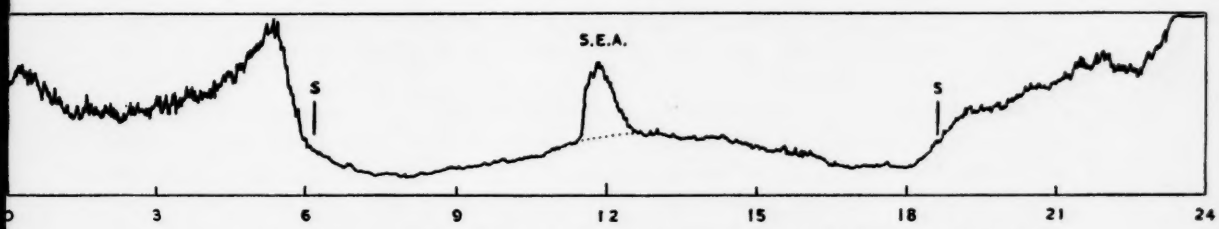


Fig. 2. A typical 24-hour record of integrated intensity of atmospherics on a frequency of 27 Kc/s, 1951 March 24. SS mark the times of local sunrise and sunset at ground level.

The atmospherics receiver was constructed for us through the kind co-operation of Mr. J. A. RATCLIFFE and members of the Radio Section, Cavendish Laboratory. It consists of a tuned aerial feeding into two tuned R.F. stages, with a cathode follower low-impedance output, which may be used, either for visual display and tuning with a C.R.O., or to feed the detector-recording circuit. The band-width of the receiver is about 200 c/s. The detector circuit introduces a time constant whose measured value is close to 60 seconds, and the rectified output is applied to a recording milliammeter. A typical 24-hour record of the integrated intensity of atmospherics on 27 Kc/s is shown in Fig. 2, and some notable examples of sudden enhancements are given in Fig. 3.

3. COMPARISON BETWEEN INDIVIDUAL FLARES AND S.E.A.S

In Fig. 4 (1–10) the continuous lines are the development curves of the flares and the broken lines those of the S.E.A.s. The points (open circles) of the S.E.A. curves were obtained as follows. Under each S.E.A. trace (see Fig. 3) a dotted line was drawn in on the original milliammeter record to represent the level of atmospherics which would have been expected in the absence of an enhancement. (This can usually be estimated with some accuracy when the complete daily record is available for inspection.) By means of a suitable graticule the ordinates above this base line were read off at one-minute intervals in the pre-maximum, and at 3 or 5 minute intervals, in the post-maximum stages. The ordinates were then scaled up or down and were plotted in Fig. 4 so as to give the same maxima as those of the flare line-widths. The zeros of the S.E.A.s were made to coincide as nearly as possible with the pre-flare and post-flare values of line-width, which usually lie between 1·0 and 2·0 A. As was shown in previous work (ELLISON, 1952),

1·5 A is the average value for the line-width in the bright hydrogen *plages* where the flares originate.

It will be seen that in five cases (3, 4, 5, 6 and 7) the S.E.A. begins at the same time as the $H\alpha$ line begins to widen: in three cases (8, 9 and 10), however, the S.E.A. commences several minutes after the flare. Nos. 1 and 2 are uncertain, since here the commencement of the flare may not have been recorded. In all 10 examples the rate of rise of the S.E.A. is *less* than that of the flare; thus the S.E.A. maxima lag behind the flare maxima by amounts which vary from $5^{m}{\cdot}0$ to $13^{m}{\cdot}5$, with an average of $8^{m}{\cdot}1$. When we reduce these figures by 1^{m} to compensate for the time constant of the receiver, we obtain an average lag of 7^{m}, in close agreement with the earlier result of 6^{m} already quoted. These facts appear to indicate, on the simplest interpretation, that the electron density at a height of about 75 km *builds up more slowly* than the number of ultraviolet light quanta incident upon the layer.

Let us next consider the decay phase. In Nos. 5, 6, 7, 8 and 9 the S.E.A. returns to zero after long delay times—15, 40, 32, 47 and 35 minutes, respectively, after the observed times of flare ending in $H\alpha$ light. With the exception of No. 6 these were all flares of the Class 2–3 category, i.e. having areas exceeding 300 millionths of the solar hemisphere (6 square degrees) and exhibiting high $H\alpha$ intensity. Although neither the end point of a flare nor of an S.E.A. can be determined with great accuracy, it is safe to conclude that after an intense flare the *D*-layer ionisation at this height may not return to normal for half an hour after the visible radiation has ceased. In the earlier results (Ellison, 1950) even longer delay times were found for Class 3^{+} flares. For the Class 1 flares, on the other hand, those of small area and low intensity, such as Nos. 2, 3 and 10, there is no appreciable interval between the two times of ending. No. 1 is an interesting case where the flare, having subsided, brightened a second time more slowly than before. The second *slow* rise did not produce any comparable effect upon the S.E.A. record. This has often been noticed; flares of the slow-increase type produce much smaller ionisation effects in *D* than those which begin with a flash.

VOL. 4 1953

4. Interpretation of Time Lag between S.E.A. Maximum and Flare Maximum

Appleton has suggested (1954) that these results indicate an interesting example of the "sluggishness" of the ionosphere. In a recent note (1953) he has given reasons for supposing that when q (the rate of electron production) passes through a maximum in any ionospheric layer, the resulting maximum of N (electron density) will be delayed by a time of the order of $1/2\alpha N$ seconds, where α is the recombination coefficient. In the case of a flare we may expect that the maximum of q is reached at the time of flare maximum and the maximum of N at the time of S.E.A. maximum. This interval (Δt) we have found to be 7 minutes, on the average.

Although it is not possible to determine N directly for the *D*-layer at a height of about 75 km, (αN) may be estimated from the equation,

$$\Delta t = \frac{1}{2\alpha N} \qquad (1)$$

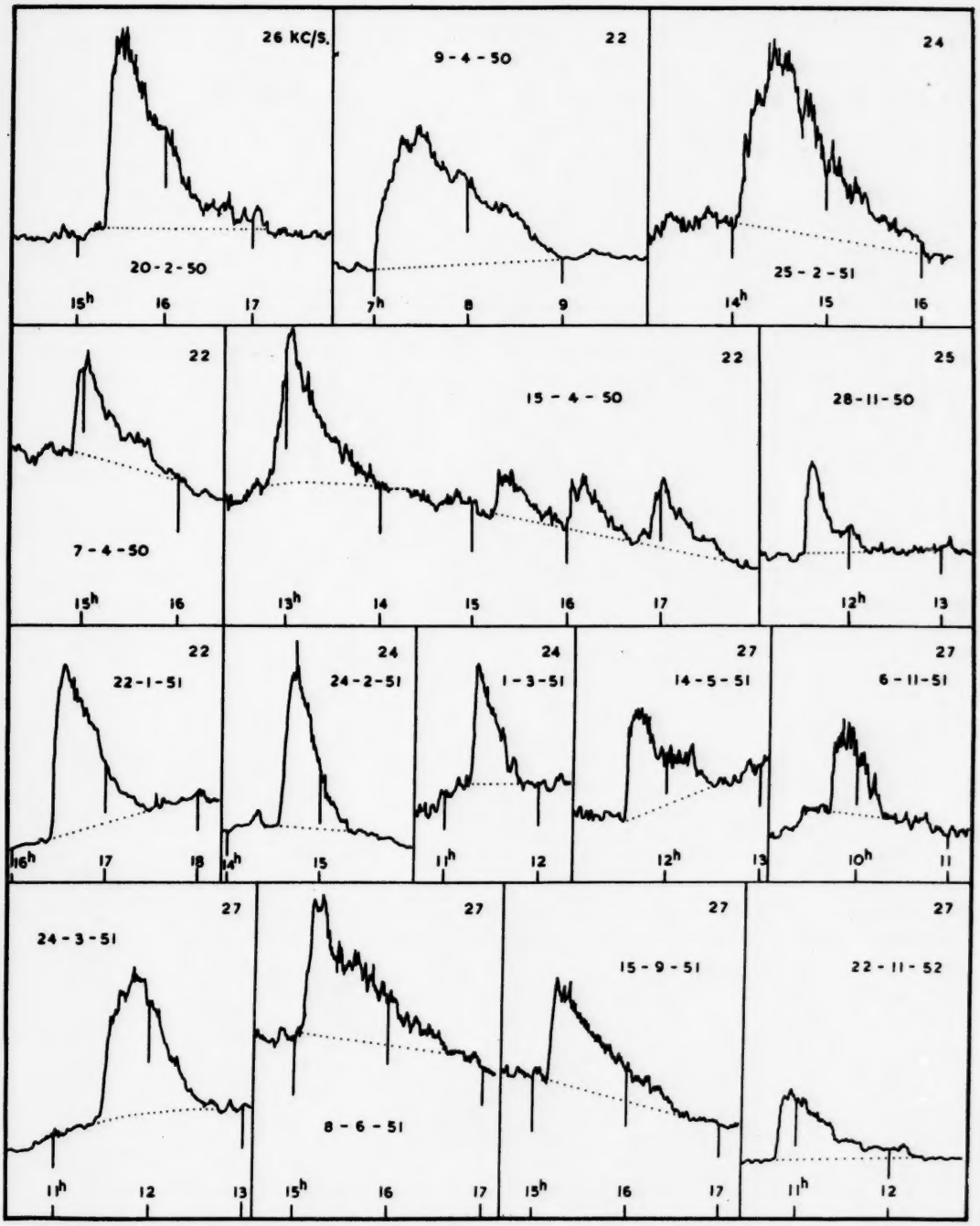


Fig. 3. Some examples of sudden enhancements of atmospherics (S.E.A.s). The figure in the top right hand corner of each diagram gives the frequency in Kc/s. The abscissae are expressed in U.T.

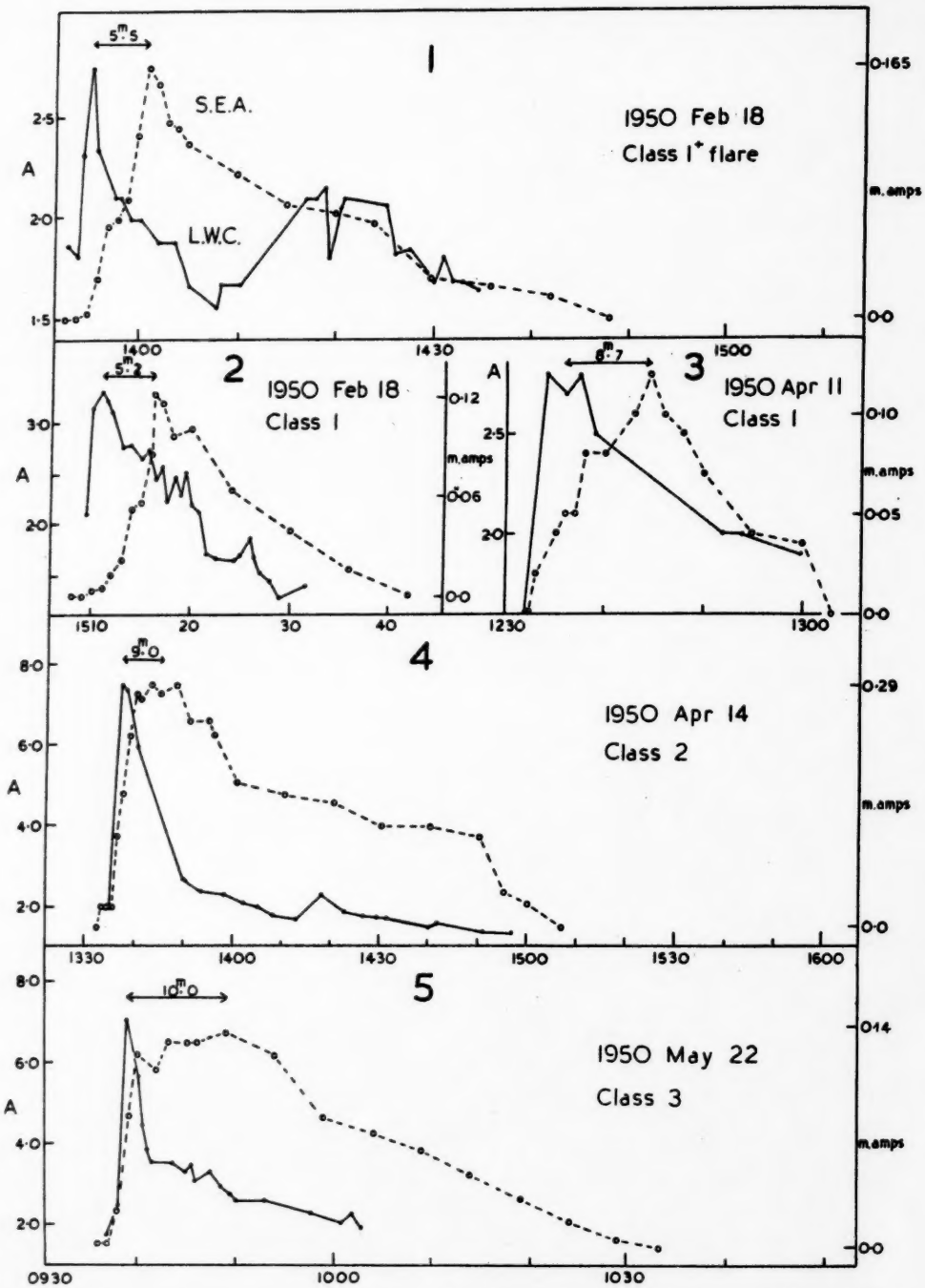
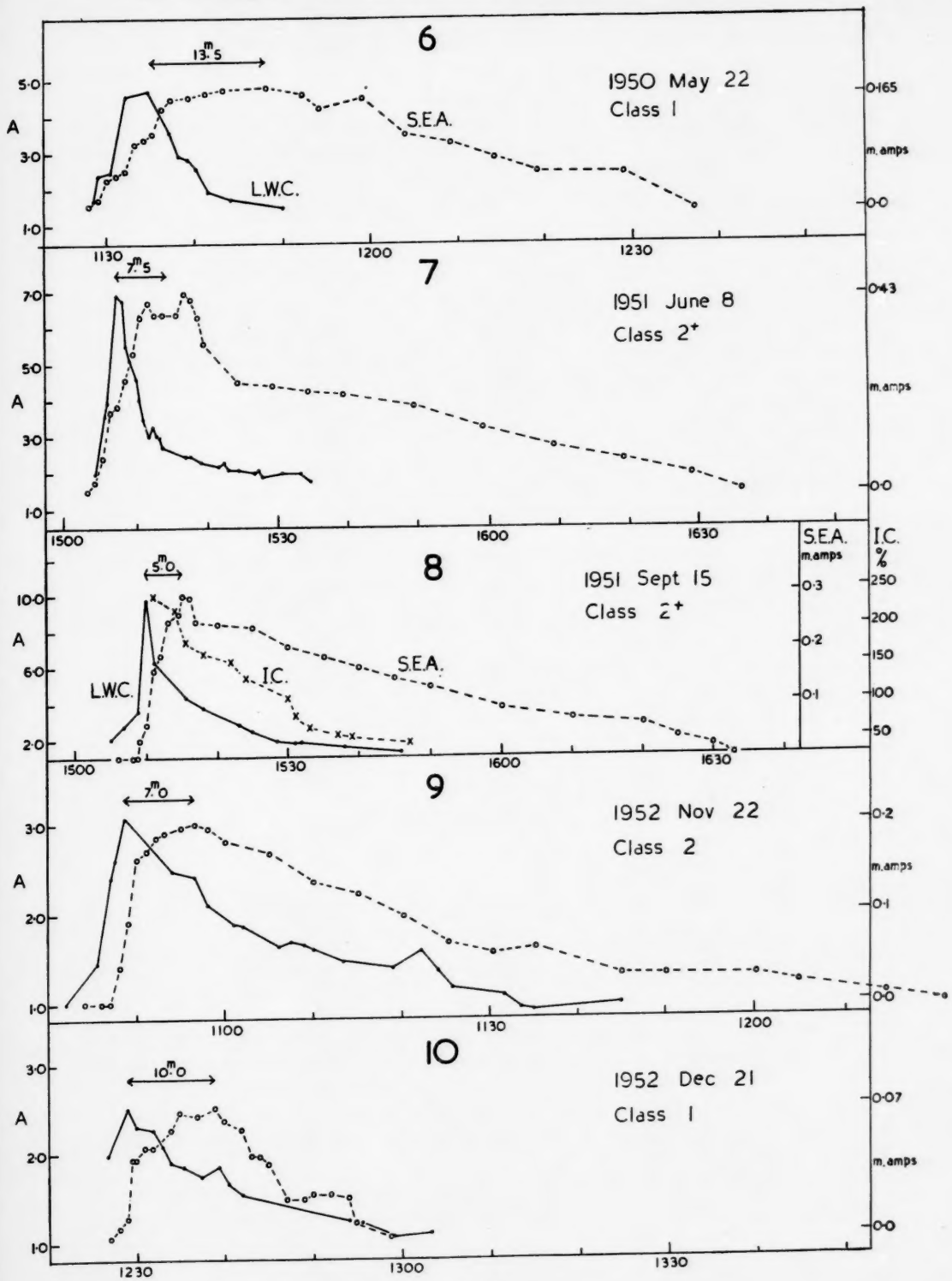


Fig. 4. The *continuous lines* are the line-width development curves for flares expressed in angstrom units (left hand ordinates). Nos. 1, 2, 5, 6 and 7 were observed at Ondřejov, and Nos. 3, 4, 8, 9 and 10 at Edinburgh. In No. 8, I.C. records the changes in central intensity of the flare $H\alpha$ line (AB of Fig. 1), expressed as a percentage of the continuous spectrum. The *broken lines* are the sudden enhancements of atmospherics (S.E.A.s) recorded at Edinburgh. The right hand ordinates give the magnitudes of the S.E.A.s in arbitrary units of milliamperes.

VOL. 4 1953


6
13.5m
1950 May 22
Class 1
S.E.A.
L.W.C.
A
5·0
3·0
1·0
0·165
m.amps
0·0
1130
1200
1230
7
7.5m
1951 June 8
Class 2+
A
7·0
5·0
3·0
1·0
0·43
m.amps
0·0
1500
1530
1600
1630
8
5.0m
1951 Sept 15
Class 2+
S.E.A.
I.C.
L.W.C.
A
10·0
6·0
2·0
S.E.A. m.amps
0·3
0·2
0·1
I.C. %
250
200
150
100
50
1500
1530
1600
1630
9
7.0m
1952 Nov 22
Class 2
A
3·0
2·0
1·0
0·2
m.amps
0·1
0·0
1100
1130
1200
10
10.0m
1952 Dec 21
Class 1
A
3·0
2·0
1·0
0·07
m.amps
0·0
1230
1300
1330

This gives for $(\alpha N)_D$ $1{\cdot}2 \times 10^{-3}$ sec^{-1} at the time of maximum of an average flare. There is another phenomenon which may be attributed to the same "sluggishness" in the response of the ionosphere to changes in the ionising radiation. This is the delay of about 30 minutes in the maximum of absorption which follows local noon. When we substitute this value for Δt in (1) we obtain for $(\alpha N)_D$ $2{\cdot}8 \times 10^{-4}$ sec^{-1}. This means that if both phenomena refer to approximately the same ionospheric level, for which we may take α to be constant, the electron density N during the flare reaches a value about 5 times the normal.

5. Rate of Onset for Flares and S.E.A.s

It is of interest to determine what is the greatest rate of onset for both flares and S.E.A.s. We have already seen that, in all cases where a detailed comparison can be made, the rate of flare rise exceeds that of the S.E.A. Now the S.E.A. records are about ten times more numerous than the flare development curves. Consequently, a selection has been made of the five most rapid S.E.A.s out of a recorded total of some 250. The early stages of these are shown in Fig. 5 (11–15). The *recorded* rate of development of an S.E.A. is naturally limited by the time-constant of the receiver. Experiments were, therefore, made to determine the response of the receiver plus recording milliammeter to an artificial S.E.A. which

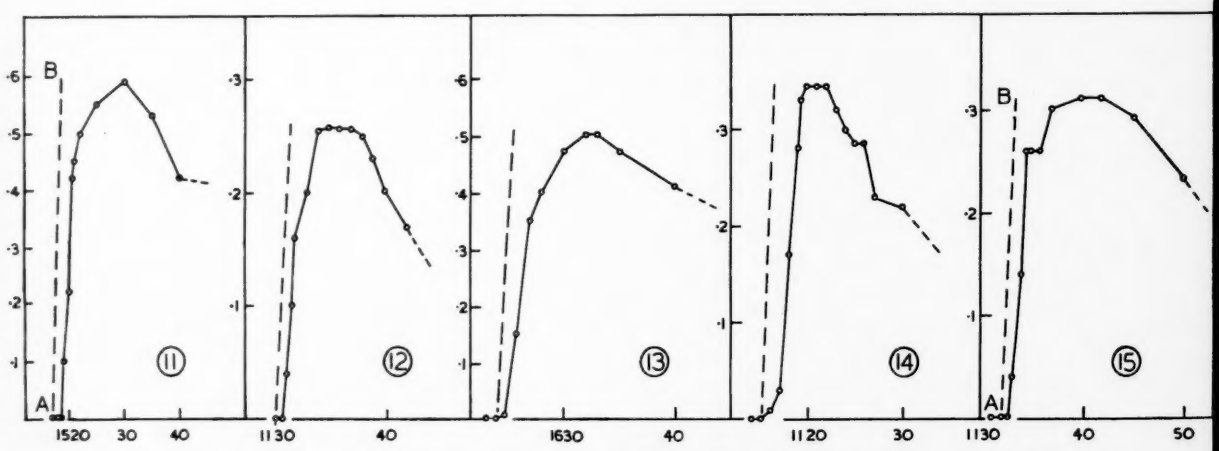


Fig. 5. S.E.A.s with unusually rapid onset. No. 11, 1950 February 20; 12, 1950 November 28; 13, 1951 January 22; 14, 1951 March 1; 15, 1951 May 14. The dashed lines AB represent the response of the receiver to a signal of comparable strength which was instantaneously switched on.

was *instantaneously* switched on. This was done by feeding into the receiver aerial, from a local 27 Kc/s oscillator, an additional signal which was adjusted so as to be comparable in strength to an average atmospherics enhancement. The signal was maintained until the milliammeter at the output end reached its maximum deflection. In Fig. 5 the rate of response of the recorder to this "instantaneous S.E.A." is given by the dashed lines AB.

Naturally, none of the observed S.E.A.s can exceed this rate of onset, but it is clear from Fig. 5 that in these specially selected cases the rise to about 60 per cent

VOL. 4 1953

of maximum is achieved at a rate which is little inferior to that which corresponds to an instantaneous "switch-on." From the 60 per cent level to maximum there is a marked slowing down, as is also noticeable in most of the records in Fig. 4. We may conclude, therefore, that these fast S.E.A.s must have been generated by exceptionally rapid ionisation and that the flashes of flare ultraviolet light responsible for them were of virtually instantaneous onset. Observations in $H\alpha$ light have suggested that such ultra-rapid flares are occasionally met with, but their progress to maximum, occurring in a time which must be considerably less than 1 minute, is too rapid to be followed by the present means of measurement.

6. Percentage Association of S.E.A.s and Flares

Since S.E.A. recorders are being increasingly used as flare indicators to supplement direct observation, it is important to know what proportion of flares of Classes 1, 2 and 3 give rise to S.E.A.s. Diurnal and seasonal changes in the ionisation of the *D* region, as well as changes in the geographical location of the main thunderstorm centres, may be expected to affect the percentage association between the two phenomena.

For a statistical analysis 413 flares were chosen during the four years for which simultaneous records of S.E.A.s were available. These flares were observed between the hours of Edinburgh sunrise and sunset, and at times when the atmospherics record was sufficiently steady to suggest that an S.E.A. could have been detected had it occurred. The flare times and classifications were taken from the *Quarterly Bulletin on Solar Activity* (*I.A.U.*) and they relate to observations made at about a dozen co-operating observatories. Where the same flare was fully observed at two or more stations and the classifications differed the lower rating was accepted. The flares were then sub-divided according to time of onset into the intervals: Sunrise–0900; 0900–1200; 1200–1500 and 1500–Sunset. These times were expressed in U.T.; since Edinburgh is in longitude 13^m west, it was considered unnecessary to convert the flare observations to local time. In Table 1, which summarises the results of the analysis, the symbol $\oplus$ denotes the occurrence of a flare which was accompanied by an S.E.A. and the symbol $\bigcirc$ a flare which was not accompanied by an S.E.A. The percentage association, that is to say the fraction *flares with S.E.A.s* divided by *total number of flares* $[\oplus/(\bigcirc + \oplus)]$, is given separately for each flare class and for each time interval.

It can be seen that for Class 1 flares (those of small area and low intensity) the fraction associated with S.E.A.s rises steadily throughout the day, from 0·22 in the early morning hours to 0·68 in the late afternoon. Disregarding the time of occurrence, 42 per cent of all Class 1 flares produce S.E.A.s. For Class 2 flares occurring after 0900, 88 per cent give rise to S.E.A.s; before 0900 the association is 62 per cent. Out of a total of 12 Class 3 flares (those of large area and high intensity) one only failed to generate an S.E.A. and this again occurred in the early morning at 0642.

The material was next divided up between the 6 summer months (March 22–September 21) and the 6 winter months (September 22–March 21), in order to ascertain if any seasonal change occurred in the percentages. Tables similar

to Table 1 were drawn up for both periods. These show that the percentage association is appreciably higher in winter than in summer. For example, between 0900 and sunset 57 per cent of Class 1 flares in winter had associated S.E.A.s, as compared with 49 per cent for the same hours in summer. In winter *all* Class 2 and Class 3 flares occurring after 0900 gave S.E.A.s. The comparisons earlier than 0900 are inconclusive owing to the small number of early morning flares observed from European stations in winter.

Table 1

Flare class / *Interval U.T.*	*Class* 1	*Class* 2	*Class* 3
Sunrise–0900	⊕ = 22, ○ = 80; P.A. = ⊕/(○ + ⊕) = 0·22	⊕ = 8, ○ = 5; P.A. = 0·62	⊕ = 1, ○ = 1; P.A. = 0·50
0900–1200	⊕ = 44, ○ = 59; P.A. = 0·43	⊕ = 15, ○ = 1; P.A. = 0·94	⊕ = 3, ○ = 0; P.A. = 1·0
1200–1500	⊕ = 41, ○ = 38; P.A. = 0·52	⊕ = 17, ○ = 2; P.A. = 0·89	⊕ = 4, ○ = 0; P.A. = 1·0
1500–Sunset	⊕ = 36, ○ = 17; P.A. = 0·68	⊕ = 13, ○ = 3; P.A. = 0·81	⊕ = 3, ○ = 0; P.A. = 1·0

The symbol ⊕ denotes a flare which was accompanied by an S.E.A.;
The symbol ○ denotes a flare which was *not* accompanied by an S.E.A.
The totals are: ⊕ + ○ = 413; ⊕ = 207; ○ = 206.

The converse problem was next considered. Are there any examples of S.E.A.s which have occurred at times when the sun was under observation in $H\alpha$ light but no flare was observed? To investigate this question each typical S.E.A. has been studied in relation to the observing times reported in the *Quarterly Bulletin*. Seven such cases have been found; and each has been examined in detail.* All seven occurred at times when a highly flare-active spot group was present on the solar disk. Three of these S.E.A.s were associated both with crochet disturbances, as reported by various magnetic stations in the sunlit hemisphere, and with short-wave radio fadeouts, as recorded by *Cable and Wireless Limited*. Of the remainder, one was associated with a crochet and one with a fadeout, leaving only two which were uncorroborated by other ionospheric records.

Now, there are no well authenticated cases of magnetic crochets, or of Dellinger-type fadeouts, having been recorded in the absence of a bright visible flare on the solar disk. Consequently, it would appear most probable that, for at least five of these seven S.E.A.s, either the observing times were reported incorrectly and the sun was not under observation, or that for some other reason the flares were missed. Thus, it may be said that there is at present no substantial evidence for the occurrence of an S.E.A. in the absence of a visible solar flare.

* The dates and times U.T. are as follows: 1950 April 7 (1453), crochet associated; April 9 (0700), crochet and fadeout associated; April 26 (1052), crochet and fadeout associated; April 27 (0934), fadeout associated: April 27 (1354); June 15 (1733), crochet and fadeout associated; 1951 June 23 (1724).

VOL. 4 1953

7. Relation between Magnitude of S.E.A. and Flare Class

The circumstances discussed in the last section would be sufficient in themselves to suggest that no simple proportionality can exist between the magnitude of an S.E.A. and the importance classification of the flare which generates it. Quite apart from the variations in sensitivity of the S.E.A. response at different hours and in different seasons, we do not yet know whether the ultraviolet light energy emitted by a flare is simply related to that of the $H\alpha$ line.

Nevertheless, it has seemed worth while to estimate the magnitudes of the *recorded* S.E.A.s and to compare them with the importances (Classes 1, 2 and 3), of the corresponding flares. The magnitude of each S.E.A. has been expressed as a percentage by determining the ratio of the S.E.A. rise in signal strength to the mean value of the signal strength before and after the S.E.A. The mean values of these magnitudes and the dispersions about the mean values are given in the following table for each flare class.

Flare class	*Mean magnitude of S.E.A.* %	*Greatest and least values* %
1	55	13–390
2	67	15–150
3	118	17–380

The figures in Column 2 indicate quite clearly that, on the average, the radiation which disturbs the ionosphere increases with flare intensity. To this extent the analysis of the S.E.A.s is in general agreement with the results obtained by workers who have examined other types of sudden ionospheric disturbance. Mention may be made of Newton and Barton's paper (1937), in which it was shown that the percentage association of radio fadings increased with flare intensity. Likewise, Bracewell and Straker (1949), in their studies of sudden phase anomalies, found that larger anomalies were associated with Class 3 flares than with those of lesser intensity.

On the other hand, the wide dispersion of the S.E.A. magnitudes about the mean values, given in Col. 3, indicates that one cannot infer from the size of an individual S.E.A. any very useful estimate of the intensity of the flare which caused it.

8. Interpretation and Conclusions

The explanation of these facts is by no means clear in the present state of our knowledge of the generation and propagation of long-wave atmospherics. There is no reason to suspect that any systematic difference exists in the estimation of flare classification at different hours of the day or in different seasons. More probably the cause lies in the real variations of signal strength received from the sources of transmission, these variations being inherent partly in the sources

themselves and partly in the ionosphere. The diurnal changes of signal strength are illustrated in Fig. 2, which is a typical 24-hour equinoctial record of atmospherics on a frequency of 27 Kc/s. The night-time level is high and fluctuating, the day level is low and steady; the main difference here is presumably caused by absence of ionospheric absorption at night. At about the time of ionospheric sunrise there is a sharp fall and the signals reach their lowest level 2–3 hours later. It is just in this period, as we have seen, that the percentage association of S.E.A.s with flares is least. However, the reduction in signal strength cannot be the sole cause of the reduced sensitivity, otherwise we should expect a similar trend before sunset. Clearly some additional factor is at work. The subsequent rise to a secondary maximum in the afternoon is believed to be due to a real increase in the number of tropical thunderstorms occurring at that hour. It will also be noted that the minimum near sunset and the rise which follows it up to the night level are neither so marked as the complementary features near sunrise.

The general level of daytime atmospherics in summer is greater than in winter by a factor of 2 or more; and this is compensated for by occasional alterations to the gain of the receiver. Here, no doubt, we have to deal with a simple effect of distance. The main transmitting centres, which are located in the tropical thunderstorm belts, may be expected to drift seasonally north and south as they follow the regions of greatest solar heating. Moreover, when the centres of transmission are furthest from us, as in winter, the received signal will have undergone, on the average, a larger number of ionospheric reflections than in summer. Consequently, any small increase in the reflection coefficient of D, such as is experienced during an S.E.A., may be expected to produce a relatively larger enhancement of the received signal in winter time. This seems to be the most probable cause of the higher percentage association in winter between minor flares and S.E.A.s, as referred to in Section 6.

The possibility has been considered whether some of these diurnal and seasonal variations of sensitivity might not be attributable to the polar diagram of the receiving aerial. This is an inverted L-antenna, 50 feet above ground level and oriented approximately NNE–SSW. At wavelengths of about 11,000 metres it seemed unlikely that such an effect could be appreciable. This conclusion was confirmed by the close correspondence found at all seasons between the daily patterns of integrated intensity of atmospherics on 27 Kc/s obtained here and those recorded by F. F. Gardner (1950) and P. W. A. Bowe (1953) at the Cavendish Laboratory, Cambridge. The latter used a vertical wire as aerial.

Flare detection is a matter of increasing importance to solar and ionospheric workers. We may summarise this aspect of the paper by saying that the S.E.A. method is of sufficient sensitivity to detect virtually all Class 2 and Class 3 flares which occur between 0900 and sunset. This means that, given a number of S.E.A. recorders well distributed in longitude, no major flare should pass undetected for lack of direct observations. It would therefore be valuable to arrange for the operation of 27 Kc/s recorders, say in America and Japan, in time for the International Geophysical Year of 1957.

The S.E.A. also enables us to fix within narrow limits the times of beginning and maximum of a flare. On the other hand, because of the many variable links,

VOL. 4 1953

already referred to, in the chain of events connecting flare and milliammeter record, the method does not enable one to distinguish with any useful degree of certainty between flares of the three different classes.

Acknowledgements—The work described was carried out at the Royal Observatory, Edinburgh. In addition to those already mentioned, the author owes much to Prof. W. M. H. Greaves for the facilities provided for this research; and to Mr. H. Seddon for his help with the maintenance of the apparatus. He also acknowledges with gratitude a grant from the Robert Cormack Committee of the Royal Society of Edinburgh for the construction of a coelostat. The coelostat, which was designed for use with the combined spectrohelioscope and spectrograph, is now on loan from the Society.

References

Appleton, E. V.	1938	Terr. Magn. **43,** 487
	1953	J. Atmosph. Terr. Phys. **3,** 282
	1954	Vistas in Astronomy, Pergamon Press London
Bowe, P. W. A.	1953	Phil. Mag. **44,** 833
Bracewell, R. N., and Straker, T. W.	1949	Mon. Not. Roy. Astron. Soc. **109,** 28
Bureau, R.	1937	Transactions of the Edinburgh Meeting (1936 September) of the International Association of Terrestrial Magnetism and Electricity, Copenhagen
		Nature **139,** 110
	1941	Notes Préliminaires du Laboratoire National de Radioélectricité Nos. 14, 52, and 119
	1947	L'Onde Electrique, No. 239 February
Dellinger, J. H.	1935	Science **82,** 351, 458
Ellison, M. A.	1950	Publications of the Royal Observatory, Edinburgh **1,** 53
	1952	Publications of the Royal Observatory Edinburgh **1,** 75
Gardner, F. F.	1950	Phil. Mag. **41,** 1259
Kleczek, J.	1952	Bulletin of the Central Astronomical Institute of Czechoslovakia **3,** 52
Link, F., and Mášková, D.	1951	Catalogue d'éruptions observées à Ondřejov de 1948 à 1950, Bulletin of the Central Astronomical Institute of Czechoslovakia **2,** 196
	1952	Bulletin of the Central Astronomical Institute of Czechoslovakia **3,** 59
Mögel, H.	1930	Telefunken-Zeitung **11,** 14
Newton, H. W.	1939	Mon. Not. Roy. Astron. Soc. **99,** 463
Newton, H. W., and Barton, H.	1937	Mon. Not. Roy. Astron. Soc. **97,** 594

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 240 to 254. Pergamon Press Ltd., London

The analysis of ionospheric records (ordinary ray) Part I

D. H. SHINN

Research Laboratories, Marconi's Wireless Telegraph Co. Ltd., Great Baddow, Essex

(*Received* 14 *May* 1953)

ABSTRACT

Tables are presented from which it is possible to construct the theoretical (h', f) curves for a "parabolic" layer (both for reflection from and for penetration through the layer) and also for a "linear" layer anywhere in the world, except at very high latitudes. The effect of the earth's magnetic field, but not that of electronic collisions, is included. Methods of deducing the electron density in the ionosphere have until now been based on calculations in which the effect of the earth's magnetic field is neglected. Necessary modifications to two of these methods are described.

INTRODUCTION

One of the most important problems in ionospheric research is the determination of the distribution of electron density as a function of height. This distribution may be deduced from observations of the apparent height, h', as a function of frequency, f, for waves vertically incident on the ionosphere. So far, the methods used for performing this deduction may be divided into two broad classes. The more accurate of the two kinds of method depends on the solution of an integral equation (APPLETON, 1930; DE GROOT, 1930; RYDBECK, 1940, 1942). But, since this kind of method generally involves some computation, another kind of method, which may be called a "comparison method," is often preferred. A recent presentation of a "comparison method" is given by RATCLIFFE (1951). In this method a simple mathematical form is assumed for the distribution of electrons, and the (h', f) curve which would result is calculated. By comparing the observed (h', f) curve with the calculated, it is then possible to decide whether the actual distribution approximates to that assumed, and, if it does, it is an easy matter to determine the fundamental parameters which characterize the distribution. The methods of BOOKER and SEATON (1940) and APPLETON and BENYON (1940) are similar to RATCLIFFE'S method in their basic ideas.

All the methods, while they are adequate for showing the major changes in the vertical distribution of electrons, cannot in general give the true distribution (even approximately), because in the theory behind the methods the effect of the earth's magnetic field is neglected. When taking into account the effect of this field a clear distinction must be drawn between the ordinary and the extraordinary rays. Here we will deal only with the ordinary ray. Both the integral equation method and the comparison method are valid for this ray only when the angle of dip of the earth's magnetic field is very small, i.e. at low magnetic latitudes. Application of either method at medium and high latitudes leads to considerable errors, as shown by SHINN and WHALE (1952). It is the purpose of this paper to present information which will enable the analysis of the ordinary reflection by RATCLIFFE'S method to be carried out at any latitude (except very near the magnetic poles). The results presented here may be applied also to the methods of BOOKER and SEATON (1940) and APPLETON and BEYNON (1940). It is hoped to

VOL. 4 1953

publish soon an account of the modifications necessary to the integral equation method in order to take into account approximately the effect of the earth's magnetic field on the ordinary ray; this has already been done for the extra-ordinary ray (at medium and high latitudes) by RYDBECK (1940).

In this paper the theoretical (h', f) curves will be presented for a "parabolic" distribution and for a "linear" distribution of electron density with height. The method of calculating these curves is the same as that used by SHINN and WHALE (5). Indeed, the results presented here are, mainly, just an extension of SHINN and WHALE'S results. The notation used is the same as theirs, namely:

$x = Ne^2/\pi m f^2$

N = number of electrons per cc

e = charge of an electron in e.s.u.

m = mass of an electron in grams

H = earth's magnetic field in oersted

f_H = gyromagnetic frequency $= He/(2\pi mc)$

c = velocity of light in free space

$y = f_H/f$

φ = angle of dip of earth's magnetic field

μ' = "group refractive index"

$$= \frac{\text{velocity of light in free space}}{\text{vertical component of group velocity of wave}}$$

1. THE (h', f) CURVE FOR REFLECTION FROM A PARABOLIC LAYER

In this section results will be presented from which it is possible to draw the theoretical (h', f) curve for reflection of the ordinary ray from a parabolic layer with any critical frequency anywhere in the world (except at very high latitudes). The layer is defined by

$$1 - N/N_0 = (1 - z/z_0)^2 \qquad (1)$$

where z = height above lower boundary of layer
z_0 = height of maximum electron density above lower boundary of layer
N_0 = maximum electron density

The effective height, h', is given by

$$h'(f) = \int_0^{z_r} \mu' \, dz + h_0$$

where h_0 = height of lower boundary of layer
and z_r = height of point of reflection of frequency f above lower boundary of layer

From equation (1)

$$dz = \tfrac{1}{2} z_0 (1 - x f^2/f_c^2)^{-\frac{1}{2}} \frac{f^2 \, dx}{f_c^2}$$

where f_c = critical frequency of layer $= \sqrt{N_0 e^2/\pi m}$

Hence
$$h'(f) = h_0 + \frac{f^2 z_0}{f_c^2} \int_0^1 \frac{\mu' t \, dt}{\sqrt{(1 - xf^2/f_c^2)}} \tag{2}$$

where $t = \sqrt{(1 - x)}$

In the absence of a magnetic field, $\mu' = 1/t$; the integration can be performed easily, and leads to

$$h'(f) = h_0 + \frac{f z_0}{2 f_c} \log_e \frac{f_c + f}{f_c - f}$$

$$= h_0 + z_0 X \left(\frac{f}{f_c}\right) \tag{3}$$

where
$$X\left(\frac{f}{f_c}\right) = \frac{f}{2f_c} \log \frac{f_c + f}{f_c - f} = \frac{f}{f_c} \tanh^{-1} \frac{f}{f_c}$$

This equation was first given by APPLETON (1937), who also showed that the parameter z_0 of the parabola which is the best approximation to a simple Chapman layer is equal to twice the scale height of the atmosphere. Equation (3) is the basis of the "comparison methods" so far used (e.g. RATCLIFFE (1951), BOOKER and SEATON (1941), APPLETON and BENYON (1940)) for deducing the height of maximum ionization and the thickness of the parabolic layer from observed (h', f) records. Such deductions are more or less in error (SHINN and WHALE (1952)), because in the theory behind equation (3) the effect of the earth's magnetic field is neglected.

In the presence of this field μ' depends also on the strength and direction of the field. Equation (2) therefore becomes

$$h'(f) = h_0 + z_0(X + Y) \tag{4}$$

where $Y = Y(f, f_c, f_H, \varphi) = \dfrac{f^2}{f_c^2} \displaystyle\int_0^1 \frac{\mu' t \, dt}{\sqrt{(1 - xf^2/f_c^2)}} - X$

Y is expressed here as a function of four variables. It may be shown that Y is dependent only on the ratios f/f_c and f_H/f_c, not on the actual values of f, f_H, and f_c. It may therefore be written $Y(f/f_c, f_H/f_c, \varphi)$.

Y is equal to zero when f/f_c is zero, when f_H/f_c is zero, or when φ is zero. In order to find $h'(f)$ for all values of f/f_c, f_H/f_c, and φ, it is necessary to compute Y for a range of values of these parameters. Since the expression for μ' is very complicated, it is quite impossible to perform the integration analytically; it was therefore done numerically, using the EDSAC electronic computer in the University Mathematical Laboratory, Cambridge; the complete computations took about ten hours machine time.

VOL. 4 1953

The results of some computations of this sort are presented by SHINN and WHALE (1952). It was estimated from these results that it would facilitate interpolation if Y were expressed as a function of X, k, and s, where

$$k = \frac{6{\cdot}4 f_H}{6{\cdot}4 f_H + f_c}$$

i.e.

$$\frac{f_c}{f_H} = 6{\cdot}4 \left(\frac{1}{k} - 1 \right)$$

and

$$s = \sec \frac{\varphi}{2}.$$

Y was computed for $X = 0{\cdot}6(0{\cdot}6)3{\cdot}0$, $s = 1{\cdot}05(0{\cdot}05)1{\cdot}3$, and $k = 0{\cdot}3(0{\cdot}1)0{\cdot}8$. The results of these computations are presented in Table 1. These are accurate to 0·2 per cent or better. The improvement in accuracy over SHINN and WHALE's results is due to more ordinates being used for the integration in the range $0 < t < \frac{1}{4}$. The practical use of Table 1 will be described later.

Notes added in proof. (i) There is an error in Tables 2 and 3 of SHINN and WHALE (1952). The quantity tabulated there is not $10^5 \frac{h' - h_0}{Z_0}$, as stated, but $10^4 \frac{h' - h_0}{Z_0}$. This error does not affect any of the rest of their paper.

(ii) E. ARGENCE and M. MAYOT (J. Geophys. Res., **58**, 1953, 147) introduce a useful and interesting approximation for μ. Using this approximation they compute $\frac{h' - h_0}{Z_0}$ as a function of f, for the location of Fribourg, and a parabolic layer with $f_c = 3{\cdot}95$ Mc/s. Their results are in good agreement with those of SHINN and WHALE.

2. THE (h', f) CURVE FOR PENETRATION THROUGH A PARABOLIC LAYER

If a pulse of frequency $f(> f_c)$ penetrates the layer and is reflected from another higher layer, then the equivalent height of reflection is increased on account of group retardation in the lower layer. For the full application of RATCLIFFE'S method it is necessary to know the group retardation in a portion of the layer. Results are presented here for the group retardation in half a layer, i.e. as far as the maximum. Numerical information from which the retardation in less than half a layer may be calculated is available from the author.

Following RATCLIFFE'S notation, the group retardation in half a layer is denoted by $\Delta h'$.

Thus

$$\Delta h'(f) = \int_0^{z_0} (\mu' - 1)\, dz$$

Use of the expression for dz deduced in the previous section, together with the substitution $x = (1 - v^2) f_c^2 / f^2$, yields

$$\Delta h'(f) = z_0 \int_0^1 (\mu' - 1)\, dv \qquad (5)$$

In the absence of a magnetic field $\mu' = (1 - x)^{-\frac{1}{2}}$; the integration can be performed easily, and leads to

$$\Delta h'(f) = z_0 \left\{ \frac{f}{2f_c} \log_e \frac{f + f_c}{f - f_c} - 1 \right\}$$

$$= z_0 \left\{ \frac{f}{f_c} \tanh^{-1} \frac{f_c}{f} - 1 \right\} \qquad (6)$$

This equation also was first given by APPLETON (1937).

Table 1. *The quantity tabulated is* $10^3 \left(\frac{h' - h_0}{z_0} - X \right) = 10^3 Y$

	φ	35·5°	49·2°	59·2°	67·1°	73·7°	79·4°	
$\frac{f_c}{f_H}$	k \ s	1·05	1·1	1·15	1·2	1·25	1·3	
14·93	0·3	17	25	30	33	37	42	
9·6	0·4	22	34	41	47	51	57	For $X = 0·6$
6·4	0·5	28	44	55	63	69	74	$\left(\frac{f}{f_c} = 0·697\right)$
4·27	0·6	34	55	70	82	90	96	
2·74	0·7	40	68	89	105	117	126	
1·6	0·8	46	82	111	134	153	166	
14·93	0·3	38	58	71	81	91	98	
9·6	0·4	50	78	98	112	124	140	For $X = 1·2$
6·4	0·5	63	100	127	148	167	182	$\left(\frac{f}{f_c} = 0·878\right)$
4·27	0·6	75	125	162	192	217	233	
2·74	0·7	89	153	204	246	278	303	
1·6	0·8	103	186	256	314	362	400	
14·93	0·3	78	123	156	182	209	222	
9·6	0·4	98	161	209	246	278	316	For $X = 1·8$
6·4	0·5	119	200	264	317	364	406	$\left(\frac{f}{f_c} = 0.955\right)$
4·27	0·6	139	242	326	400	463	511	
2·74	0·7	160	289	400	498	581	649	
1·6	0·8	181	340	485	618	736	837	
14·93	0·3	146	247	328	398	469	506	
9·6	0·4	176	307	419	516	604	700	For $X = 2·4$
6·4	0·5	204	336	508	641	756	881	$\left(\frac{f}{f_c} = 0.985\right)$
4·27	0·6	231	426	602	777	932	1080	
2·74	0·7	258	489	712	929	1137	1330	
1·6	0·8	283	556	829	1106	1384	1654	
14·93	0·3	244	442	625	798	982	1103	
9·6	0·4	281	524	758	984	1215	1466	
6·4	0·5	314	600	884	1170	1461	1790	For $X = 3·0$
4·27	0·6	345	673	1009	1360	1727	2124	$\left(\frac{f}{f_c} = 0·995\right)$
2·74	0·7	375	746	1138	1562	2022	2519	
1·6	0·8	402	822	1277	1785	2357	3002	

VOL. 4 1953

Table 2. The quantity tabulated is $\left(Z - \frac{\Delta h'(f)}{z_0}\right) 10^3$

k \ s	1·05	1·1	1·15	1·2	1·25	1·3	
0·4	27	37	43	47	49	50	For $Z = 0·3$ $\left(\frac{f}{f_c} = 1·330\right)$
0·5	35	51	59	65	69	71	
0·6	45	67	80	88	93	96	
0·7	53	86	106	118	126	130	
0·8	55	105	137	156	168	175	
0·4	60	97	119	133	141	146	For $Z = 0·6$ $\left(\frac{f}{f_c} = 1·123\right)$
0·5	68	120	154	175	189	197	
0·6	65	139	191	222	243	256	
0·7	49	145	222	272	304	324	
0·8	18	120	230	315	369	402	
0·4	62	146	209	248	272	286	For $Z = 0·9$ $\left(\frac{f}{f_c} = 1·056\right)$
0·5	47	146	237	302	342	366	
0·6	20	125	248	348	414	451	
0·7	−18	79	225	368	477	540	
0·8	−67	−1	146	332	511	624	
0·4	14	138	263	355	418	458	For $Z = 1·2$ $\left(\frac{f}{f_c} = 1·028\right)$
0·5	−33	83	243	386	487	554	
0·6	−80	−3	176	379	539	644	
0·7	−124	−110	54	303	546	718	
0·8	−163	−224	−133	117	465	752	
0·4	−78	14	196	391	542	641	For $Z = 1·5$ $\left(\frac{f}{f_c} = 1.015\right)$
0·5	−129	−99	83	335	562	724	
0·6	−175	−215	−83	191	522	782	
0·7	−214	−326	−288	−44	379	781	
0·8	−249	−431	−513	−370	68	668	

In the presence of a magnetic field μ' depends also on the strength and direction of the field. It is convenient to express $\Delta h'(f)/z_0$ as a function of k, s, and Z, where $Z = (f/f_c)\tanh^{-1}(f_c/f) - 1$, and k and s are as defined in the previous section. $\Delta h'(f)/z_0$ was computed for $Z = 0·3(0·3)1·5$, $s = 1·05(0·05)1·3$, and $k = 0·4(0·1)0·8$. The results of these computations are presented in Table 2. The practical use of Table 2 will be described in the next section.

3. Use of the Tables in Constructing Test Scales (Ratcliffe's Method)

In the method of analysis used by Ratcliffe, various theoretical (h', f) curves are reproduced on a transparent material. These are laid over the observed curve, and the best fitting curve is chosen. Ratcliffe used curves based on theory in which the effect of the earth's magnetic field is neglected. From the information in Tables 1 and 2 it is possible to draw similar curves with the effect

of the earth's magnetic field taken into account. Use of these curves will, of course, yield a more accurate method of analysis.

The shape of the (h', f) curve depends on the angle of dip and on the strength of the earth's magnetic field. These elements are presented in Table 3 for several magnetic observatories. This data was obtained from a comprehensive survey of geomagnetic information given by JOHNSTON (1951), together with the assumption that the field varies with height as $(R + h)^{-3}$, where R is the radius of the earth, taken to be 6370 Km.

Table 3 (*mean values estimated for* 1953)

Observatory	$\sec\frac{\varphi}{2}$	f_H (*in Mc/s*) *at* 120 *Km*	f_H (*in Mc/s*) *at* 300 *Km*
Amberley, New Zealand	1·207	1·574	1·450
Watheroo, Australia	1·182	1·529	1·409
Hermanus, S. Africa	1·182	0·826	0·761
Abinger, England	1·197	1·247	1·149
Eskdalemuir, Scotland	1·220	1·276	1·175
Cheltenham, Maryland, U.S.A.	1·230	1·506	1·387
Lerwick, Shetland Islands	1·244	1·303	1·201
Agincourt, Ontario, Canada	1·257	1·548	1·426

For Abinger sec $\varphi/2$ may be taken to be 1·20. Table 4 shows values of h' for Abinger deduced from Table 1, while Table 5 shows values of $\Delta h'$ for Abinger deduced from Table 2. These values are used to plot the curves shown in Fig. 1. These curves are similar to those used by RATCLIFFE in his analysis; further curves for other thicknesses of layer can easily be constructed if required. The value of f_H at a height of 300 km has been used. Strictly speaking, the correct value to take, according to SHINN and WHALE (1952), is that at a height of $h_0 + 0{\cdot}65z_0$. Errors due to taking the wrong height or the wrong critical frequency will be discussed later.

In order to construct Tables 4 and 5 it was necessary to interpolate for X and k in the columns of Tables 1 and 2 headed sec $\varphi/2 = 1{\cdot}2$. For another location it may be necessary to interpolate also for sec $\varphi/2$. For example, at Cheltenham, Maryland, sec $\varphi/2 = 1{\cdot}230$. The results of interpolating for sec $\varphi/2$ in Table 1 are shown in Table 6. Further interpolation in Table 6 would be necessary in order to produce a table like Table 4, from which test scales can be constructed. The information in Tables 1 and 2 may be used in a similar way for the construction

VOL. 4 1953

of test scales for any location in the world (at magnetic latitudes up to about 70°, or angles of dip up to about 80°), provided only that the magnetic elements are known.

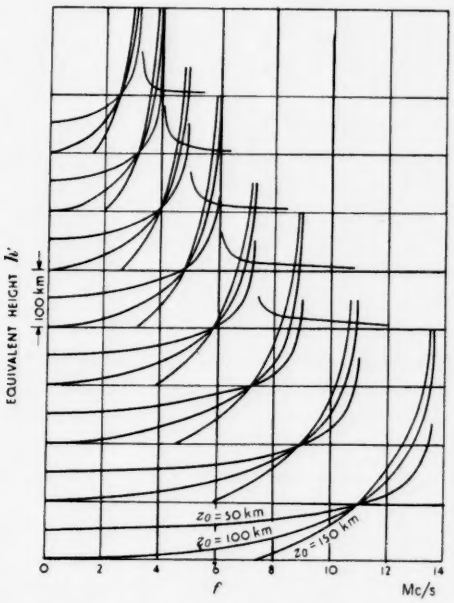


Fig. 1. The theoretical (h', f) curves for parabolic layers with critical frequencies 3·15, 3·96, 4·90, 6·02, 7·35, 8·99, 11·03, and 13·66 Mc/s, above S.E. England. The curves are for semi-thicknesses of layer of 50, 100, and 150 Km. The curves for penetration are for a semi-thickness of 50 Km only.

Table 4. *The quantity tabulated is* $10^3(h' - h_0)/z_0 = 10^3(X + Y)$, *when sec* $(\varphi/2) = 1{\cdot}20$

	k	0·35	0·4	0·45	0·5	0·55	0·6	0·65	0·7	0·75
X	f_c / f/f_c	13·66	11·03	8·99	7·35	6·02	4·90	3·96	3·15	2·45
0·3	0·520	321	324	329	333	337	342	347	353	358
0·6	0·697	640	647	655	663	672	682	693	705	718
0·9	0·806	963	974	986	999	1013	1029	1047	1066	1088
1·2	0·878	1296	1312	1329	1348	1369	1392	1418	1446	1478
1·5	0·925	1642	1666	1691	1718	1747	1779	1815	1854	1897
1·8	0·955	2013	2046	2080	2117	2157	2200	2247	2298	2354
2·1	0·974	2415	2460	2507	2556	2609	2665	2725	2790	2860
2·4	0·985	2857	2916	2977	3041	3107	3177	3250	3329	3414
2·7	0·991	3346	3421	3498	3577	3657	3740	3827	3919	4017
3·0	0·995	3892	3894	4077	4170	4264	4360	4459	4562	4670

Table 5. The quantity tabulated is $10^3\Delta h'/z_0$, *when sec* $(\varphi/2) = 1{\cdot}20$

f/f_c \ f_c	7·35	6·02	4·90	3·96	3·15
1·689	125	120	115	109	101
1·330	235	224	212	198	182
1·193	334	316	297	276	254
1·123	425	402	378	353	328
1·082	510	484	459	435	416
1·056	598	573	552	537	532
1·040	695	677	667	669	687
1·028	814	811	821	848	897
1·020	967	990	1030	1091	1178
1·015	1165	1226	1309	1415	1544

Table 6. The quantity tabulated is 10^3Y, *when sec* $(\varphi/2) = 1{\cdot}23$

X \ k	0·3	0·4	0·5	0·6	0·7	0·8
0·6	35	49	67	87	113	146
1·2	87	119	160	208	266	344
1·8	199	265	346	439	550	691
2·4	443	569	710	872	1055	1273
3·0	912	1121	1342	1577	1834	2120

VOL. 4 1953

4. Discussion of the Results for a Parabolic Layer

In Figs. 2a and 2b h' and $\Delta h'$ are plotted against f/f_c for an angle of dip 67·1° (i.e. sec $\varphi/2 = 1{\cdot}2$), and various values of f_H/f_c. f_H is taken to be 1·149 Mc/s. For values of f/f_c less than unity, h' exceeds Xz_0, i.e. the net effect of the magnetic field is to cause a delay. For values of f/f_c greater than unity, $\Delta h'$ is sometimes greater than and sometimes less than Zz_0; the effect of the magnetic field is thus sometimes to increase and sometimes to decrease the group retardation of a wave which penetrates through a parabolic layer.

This effect can readily be explained by reference to Fig. 15 of SHINN and WHALE. This figure shows that, when f_c^2/f^2 is less than about 0·92, the effect of the magnetic field is to decrease the group refractive index at all heights in the layer; therefore, for frequencies greater than about $1{\cdot}04\,f_c$, the retardation in the layer is less than it would be in the absence of a magnetic field. For larger values of f_c^2/f^2, the group refractive index for small electron densities is less than its value in the absence of a magnetic field, but for the large electron densities near the maximum of the layer it is greater than its value in the absence of a magnetic field. As f_c/f approaches unity, this causes sufficient retardation to make the total retardation greater than that in the absence of a magnetic field. This effect is more marked for low critical frequencies than for high critical frequencies.

The curves shown in Fig. 2a are similar in shape to those often observed in practice (e.g. Fig. 4b of SHINN and WHALE).

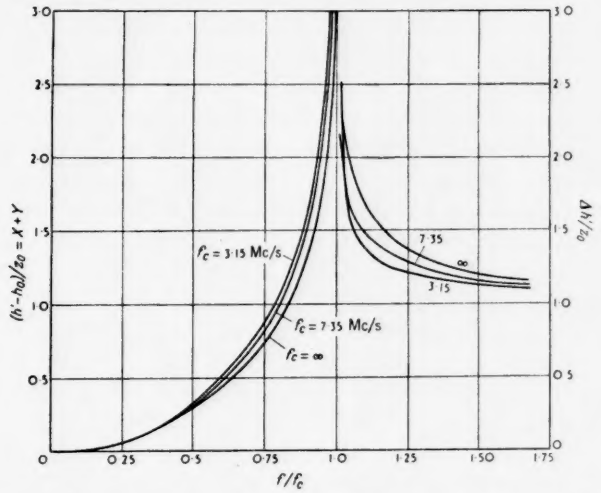


Fig. 2a. Normalized theoretical (h', f) curves for a parabolic layer above S.E. England.

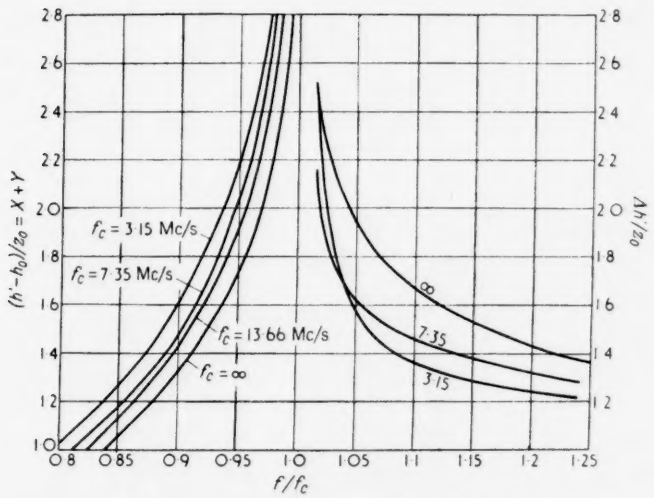


Fig. 2b. An enlarged version of part of Fig. 2a.

5. ACCURACY OF THE METHOD OF ANALYSIS FOR A PARABOLIC LAYER

Four fundamental assumptions have been made:

(i) the ionosphere is horizontally stratified,

(ii) the distribution of electron density is "parabolic,"

(iii) the ray theory of propagation is valid,

(iv) the effect of electronic collisions may be neglected.

Let us now examine the status of these assumptions.

The first assumption conflicts with the well-known fact that the ionosphere has considerable irregularities in the horizontal plane. However, these irregularities generally have very little influence on the time delay of the reflected signal; under such circumstances the distribution of electron density deduced from measurements of time delay must be regarded as the mean density of electrons at a particular height. Occasionally the irregularities may be so large as to cause the signal to be reflected at a height considerably lower than that for which the mean electron density is sufficient to reflect the wave; for reflections of this sort (i.e. from "clouds" of ionization, usually in the *E* Region), the analysis is not applicable.

The second assumption is often nearly true in practice, at least over a limited region in the ionosphere. The third assumption is adequate. But the fourth assumption may not hold for the *E* Region. These last two assumptions are dealt with in detail by RYDBECK (1944).

There still remains the question of the intervals of s and f_c to be used in the construction of test scales. A suitable criterion for this is the allowable error in the computed value of $h'(f)$. If the true critical frequency is f_c, and the critical frequency of the theoretical curve with which it is compared is $f_c + \delta f_c$, it may be shown that the error in f/f_c is given by $\delta(f/f_c) = (1 - f/f_c)\delta f_c/f_c$, provided that a linear scale of frequency is used. Hence, allowing for all possible errors, it follows from equation (4) that

$$\frac{\delta h'(f)}{z_0} = \left(1 + \frac{\partial Y}{\partial X}\right) \frac{dX}{d(f/f_c)} \frac{\delta f_c}{f_c} \left(1 - \frac{f}{f_c}\right) + \frac{\partial Y}{\partial k} \left(\frac{\partial k}{\partial f_c} \delta f_c + \frac{\partial k}{\partial f_H} \delta f_H\right) + \frac{\partial Y}{\partial s} \delta s$$

Now $$\frac{dX}{d(f/f_c)}\left(1 - \frac{f}{f_c}\right) = \frac{n}{1+n} + (1 - n) \tanh^{-1} n,$$

where $n = f/f_c$

$$= p(f/f_c), \text{ say.}$$

(It may be shown that $0{\cdot}5 < p < 0{\cdot}67$ for all values of f/f_c in the range $0{\cdot}35 < f/f_c < 1{\cdot}0$.)

Also $$f_c \frac{\partial k}{\partial f_c} = -k(1 - k) \text{ and } f_H \frac{\partial k}{\partial f_H} = k(1 - k)$$

Hence $$\frac{\delta h'(f)}{z_0} = \left[p(f/f_c)\left(1 + \frac{\partial Y}{\partial X}\right) - k(1 - k) \frac{\partial Y}{\partial k}\right] \frac{\delta f_c}{f_c} + k(1 - k) \frac{\partial Y}{\partial k} \frac{\delta f_H}{f_H} + \frac{\partial Y}{\partial s} \delta s \qquad (7)$$

In order to estimate the error in a particular case in S.E. England, we must refer to Table 4. Let us consider a critical frequency in the region of 7·4 Mc/s,

VOL. 4 1953

and $f/f_c = 0{\cdot}974$, i.e. $k = 0{\cdot}5$ and $X = 2{\cdot}1$. At this point $\partial Y/\partial X \simeq 0{\cdot}54$, $\partial Y/\partial k \simeq 1{\cdot}0$, $\partial Y/\partial s \simeq 1{\cdot}6$, and $p(f/f_c) \simeq 0{\cdot}55$. It follows that

$$\delta h'(f) \simeq z_0[0{\cdot}60\,\delta f_c/f_c + 0{\cdot}25\delta f_H/f_H + 1{\cdot}6\delta s]$$

When $\dfrac{f}{f_c} = 0{\cdot}925$, $\dfrac{\delta h'(f)}{z_0} = 0{\cdot}65\delta f_c/f_c + 0{\cdot}12\delta f_H/f_H + 0{\cdot}6\delta s$

When $\dfrac{f}{f_c} = 0{\cdot}991$, $\dfrac{\delta h'(f)}{z_0} = 0{\cdot}58\delta f_c/f_c + 0{\cdot}40\delta f_H/f_H + 3{\cdot}8\delta s$

Furthermore, if f_H is chosen to be the gyrofrequency at a height of 300 Km, then $\delta f_H/f_H = -\delta h_1/2200$, where $h_1 = h_0 + 0.65z_0$, measured in kilometres, and $\delta h_1 = h_1 - 300$. The largest numerical value of h_1 likely to be encountered is -180.

It appears from these figures that the error introduced in h' by the use of Fig. 1 for the analysis of curves in England is less than $0{\cdot}07z_0$ for a critical frequency in the region of 7·4 Mc/s. The error is less for greater critical frequencies, and more for smaller critical frequencies. A similar analysis shows that for a critical frequency in the region of 3 Mc/s the error is less than $0{\cdot}10z_0$. If the error in the theoretical value of h' is more than can be tolerated, it is easy to reduce it, by reducing δf_c (i.e. by plotting curves as in Fig. 1 for a closer spacing of values of f_c), δf_H (i.e. by constructing a series of curves as in Fig. 1 for two or three values of f_H), or δs (i.e. by computing a table like Table 4 for $s = 1{\cdot}197$), as required. An alternative criterion for the allowable error is the error in f. This is given by

$$\delta f = \frac{\delta h'}{z_0}\,\frac{f_c - f}{p(1 + \partial Y/\partial X)}$$

6. Modifications to Appleton and Beynon's Method for a Parabolic Layer

In this method the observed value of h' is plotted against $X(f/f_c)$. It is evident from equation (3) that in the absence of a magnetic field the theoretical (h', X) curve is a straight line, and that the slope of this line and its intercept on the h' axis yield the values of z_0 and h_0 respectively. It is also clear from equation (4) that in the presence of the field the theoretical $(h', X + Y)$ curve is a straight line, and that its slope and intercept yield the values of z_0 and h_0. Therefore the observed values of h' should be plotted not against X but against $X + Y$. The value of $X + Y$ depends on f_c, φ, and f/f_c; it is tabulated for Abinger in Table 4. An application of this method to some observed (h', f) curves is given by Shinn and Whale (1952).

The error in h' due to choosing the wrong value of f_c, s, or f_H is given by

$$\frac{\delta h'(f)}{z_0} = \frac{\partial Y}{\partial k}\left(\frac{\partial k}{\partial f_c}\,\delta f_c + \frac{\partial k}{\partial f_H}\,\delta f_H\right) + \frac{\partial Y}{\partial s}\,\delta s$$

$$= k(1 - k)\,\frac{\partial Y}{\partial k}\left(\frac{\delta f_H}{f_H} - \frac{\delta f_c}{f_c}\right) + \frac{\partial Y}{\partial s}\,\delta s$$

For the same conditions as in the previous section, and for $f/f_c = 0{\cdot}974$,

$$\delta h' = z_0[0{\cdot}25\delta f_H/f_H - 0{\cdot}25\delta f_c/f_c + 1{\cdot}6\delta s]$$

If the values in Table 4 are used, the maximum error in h' when $f/f_c = 0{\cdot}974$ is about $0{\cdot}05z_0$. The error may be easily reduced by using a more nearly correct value of f_H, and interpolating for k in Table 4; for extreme accuracy it may be necessary to interpolate also for s in Table 1.

7. (h', f) Curves for Reflection from a Linear Layer

The layer is defined by

$$Ne^2/\pi m = f_0^2 = \alpha z$$

where α is a constant. It may be shown that

$$h'(f) - h_0 = \frac{2f^2}{\alpha}\int_0^1 \mu' t\, dt \tag{8}$$

z, h_0, and t are defined in section 1. It is convenient to use the quantity $H'(f)$, defined by

$$H'(f) = \frac{(h'(f) - h_0)\alpha}{2f^2}$$

Thus $H'(f)$ is the ratio between $(h'(f) - h_0)$ in the presence of the magnetic field and $(h'(f) - h_0)$ in the absence of a magnetic field. It follows from equation (8) that

$$H'(f) = \int_0^1 \mu' t\, dt$$

Now $\mu' t$ is a function of y and φ. Therefore $H'(f)$ is a function of y and φ. In Fig. 3 $H'(f)$ is plotted against φ for various values of $1/y = f/f_H$. The way to use this information in interpreting reflection from linear layers is obvious. It is noteworthy that, as for the parabolic layer, the magnetic field always causes a retardation of the ordinary ray. This is in accordance with the calculations of GOUBAU (1934) and POEVERLEIN (1951). GOUBAU's results are not presented sufficiently accurately for comparison. POEVERLEIN found that H′ is equal to 1·112 and 1·136 for $1/y = 2{\cdot}675$ and $\varphi = 65°$ and 80° respectively. The corresponding values from Fig. 3 are about 1·114 and 1·141 respectively. The differences between these figures are probably due to errors in integration or interpolation; they are too small to be of practical significance.

In Fig. 3 the curve for $f = 0$ is drawn. This curve is of no intrinsic interest, since the ray theory is not valid for very low frequencies, but it shows the limiting value given by the ray theory as y tends to infinity, and is thus useful for interpolation.

Conclusions

Until now analysis of (h', f) curves in order to deduce the distribution of electrons in the ionosphere has been based on theory in which the effect of the earth's magnetic field is neglected. Information presented here enables the effect of the

VOL. 4 1953

field to be taken into account when using the method of RATCLIFFE (1951), which is based on that of BOOKER and SEATON (1940), or when using the method of APPLETON and BEYNON (1940). The integral equation method (APPLETON, 1930; DE GROOT, 1930; RYDBECK, 1940, 1942) can also be modified in order to take into account approximately the effect of the field; it is hoped to publish information about this modification soon.

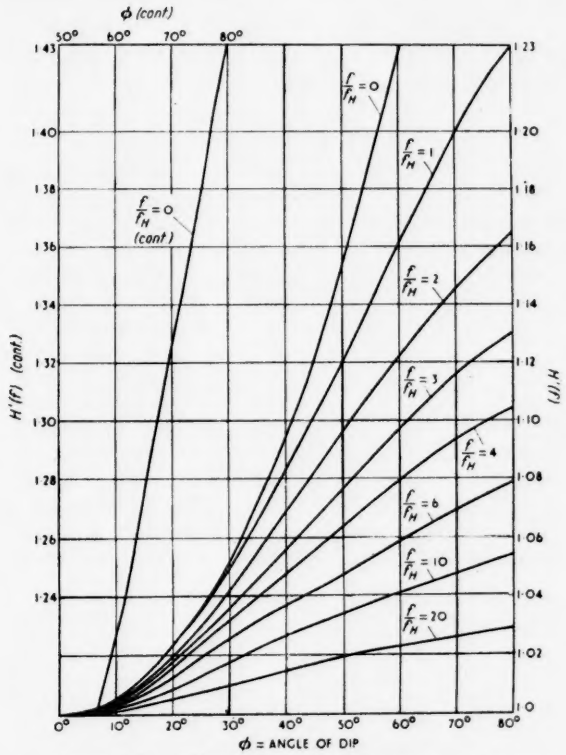


Fig. 3. Curves for calculating the effective height of a linear layer.

$$H'(f) = \frac{(h' - h_0) \text{ in the presence of a magnetic field}}{(h' - h_0) \text{ in the absence of a magnetic field}}$$

The analysis presented here applies to the ordinary ray. The theory neglecting the effect of the earth's magnetic field is valid for this ray only very near the magnetic equator. The information presented here extends the theoretical results, for reflection of the ordinary ray from a parabolic layer or from a linear layer, up to magnetic latitudes of about 70° (angles of dip up to about 80°). For the extraordinary ray a method of analysis which is valid at high latitudes, greater than about 45°, has been used by RYDBECK (1940).

The results presented here show that the earth's magnetic field always increases the group delay for the wave reflected from a linear layer or a parabolic layer. For a wave which penetrates a parabolic layer, the magnetic field increases the group delay for frequencies just greater than the critical frequency, but decreases

the group delay for higher frequencies. This effect has been noticed on observed (h', f) curves.

Acknowledgements—This paper is published by permission of *Marconi's Wireless Telegraph Company, Ltd.* The computations were carried out on the EDSAC at the University Mathematical Laboratory, Cambridge, through the courtesy of the Director, Dr. M. V. WILKES. I am much indebted to Dr. WILKES and his staff, and especially to Dr. K. G. BUDDEN of the Cavendish Laboratory, for their assistance in the computation. I am also grateful for the advice and encouragement of Mr. J. A. RATCLIFFE, of the Cavendish Laboratory, and Mr. G. MILLINGTON, of *Marconi's Wireless Telegraph Company.*

REFERENCES

APPLETON, E. V.	1930	Proc. Phys. Soc. **42**, 321
	1937	Proc. Roy. Soc. A **162**, 451
APPLETON, E. V. and BEYNON, W. J. G.	1940	Proc. Phys. Soc. **52**, 518
BOOKER, H. G. and SEATON, S. L.	1940	Phys. Rev. **57**, 87
DE GROOT, W.	1930	Phil. Mag. **10**, 521
GOUBAU, G.	1934	Hochfrequenz u. Elektroakust. **44** 17 and **44**, 138
JOHNSTON, H. F.	1951	J. Geophys. Res. **56**, 431
POEVERLEIN, H.	1951	Z. angew. Phys. **3**, 135
RATCLIFFE, J. A.	1951	J. Geophys. Res. **56**, 463
RYDBECK, O. E. H.	1942	Trans. Chalmers Univ. Tech. (Gothenburg), Nr. 3
	1940	Phil. Mag., **30** 282
	1944	Trans. Chalmers Univ. Tech. (Gothenburg), Nr. 34, p. 139
SHINN, D. H. and WHALE, H. A.	1952	J. Atmos. Terr. Phys. **2**, 85

VOL. 4 195

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 255 to 270. Pergamon Press Ltd., London

Measurements of winds in the upper atmosphere by means of drifting meteor trails I

D. S. Robertson, D. T. Liddy and W. G. Elford

Department of Physics, University of Adelaide, South Australia

(*Received* 10 *June* 1953)

ABSTRACT

The paper describes some new radio techniques for the measurement of upper atmosphere winds by means of drifting meteor trails. A continuously recording doppler-radar system operating on 27 mc is described and some typical records are analysed. An accompanying paper deals with the experimental results.

Introduction

Manning, Villard and Peterson (1950) have described a radio method of measuring upper atmosphere winds from the drift of meteor trails. The purpose of the present paper is to describe a continuously recording method for measuring winds by observing the drift of meteor trails. This paper is concerned with techniques, and the discussion of results is reserved for another paper.

The first part of the paper deals with techniques for measuring the drift velocities and space co-ordinates of echoing centres of meteor trails. A simple method of measuring winds without regard to their height is also described.

The second part of the paper is an account of the equipment in use at Adelaide. The equipment comprises a transmitter operating on a frequency of 27 mc, located at Adelaide, and receivers and a recorder near Salisbury, 19·6 km north of Adelaide. The reception of a meteor echo at Salisbury starts the recorder which records the line of sight component of drift velocity of the point of reflection, its space co-ordinates and the time at which the echo occurs. The apparatus runs continuously for two or three weeks, in which time between 700 and 1,000 useful echoes are usually recorded. From an analysis of these it is possible to calculate the average diurnal behaviour of the wind at heights between 80 and 110 kilometres. During periods of great activity it is also possible to observe the wind almost continuously.

The third section of the paper deals with the interpretation of the records, and the final section with calibration of the direction-finding equipment.

Long enduring echoes of the type used by Greenhow (1952) in his wind gradient measurements, showing severe fading, have not been used. Although many have been recorded, very few of these confused echoes can be resolved into two echoing centres using a doppler technique.

1. Techniques

1.1 *Drift velocity*

The line of sight component of drift velocity is found by beating the reflected signal with the direct signal and measuring the frequency of the resulting "doppler beat."

All recording is done by photographing cathode ray tubes on horizontally

moving bromide paper, the doppler signal being applied to the vertical plates of a cathode ray tube. A typical record is shown in Fig. 1. The cusped form of the record arises from the fact that the amplitude of the ground wave is adjusted to be approximately equal to that of an average echo. The record of Fig. 1 is produced by a normal receiver, whose voltage output is proportional to the field strength of the electric field at the aerial, the resultant field being the vector sum of the fields of the reflected and direct waves. The amplitude of the ground wave is adjusted by choosing an appropriate height for the receiver aerial. The transmitted power is beamed upwards in a cone of radiation with a semi angle of approximately 40°, but there is enough direct signal to provide the required ground wave at Salisbury. The line of sight component v_r of the drift velocity is

$$v_r = f\frac{\lambda}{2} \text{ metres/sec} \tag{1}$$

where f = the doppler shift in cycles/sec.

and λ = the radiofrequency wavelength in metres.

In our case, $\lambda = 11{\cdot}2$ metres, so that

$$v_r = 5{\cdot}6.f \text{ metres/sec.}$$

1.2 *Doppler sense determination*

It will be seen from Fig. 1 that with the doppler signal there are associated "spikes" whose purpose is to indicate whether the doppler shift is positive or negative. A positive shift denotes an approaching echo and a negative shift a receding echo.

These sense spikes are derived as follows:

Consider the vector sum of the ground wave and reflected wave, as shown in Fig. 2a, where G_1 is the ground wave, S is the reflected wave and R_1 is their vector sum.

Fig. 1. A typical doppler record.

Let S rotate in a positive sense relative to G_1, so that S represents a signal with a positive doppler shift, i.e. an approaching echo. R_1 is growing in amplitude and approaching its maximum value.

Next consider Fig. 2b. Here, the ground wave, G_2 has been retarded in phase by 90° relative to G_1, and the resultant R_2 has passed its maximum amplitude and is shrinking. Thus R_2 leads R_1 in phase.

Now let S_1 be a receding echo. Its sense of rotation is reversed and R_1 will have passed its maximum, whereas R_2 will be approaching its maximum value. Thus R_1 will lead R_2 in phase.

In Fig. 1. the full line represents R_1, and the peaks of the spikes represent R_2. These spikes are produced by suddenly retarding the phase of the transmitted wave by $\pi/2$. The phase retarded ground wave reaches Salisbury about half a

VOL. 4 1953

millisecond before the phase-retarded reflected wave, so that during this brief period the phase of the ground wave is retarded by $\pi/2$ relative to its original value, but the phase of the reflected wave is unchanged, and the conditions of Fig. 2b prevail. Condition 2a is restored upon the arrival of the phase shifted echo, and the receiver output then returns to its original level.

This sudden change in the phase of the transmitted signal occurs 50 times a second. After the sudden change, the phase drifts slowly back to its original value. The waveform of the phase modulation applied to the transmitter is shown in Fig. 3. Thus it will be seen that if the phase of the wave traced out by the tops

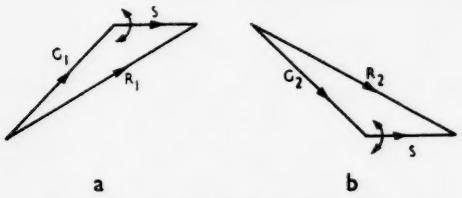


Fig. 2. Vector diagrams for demonstrating the method of sense determination.

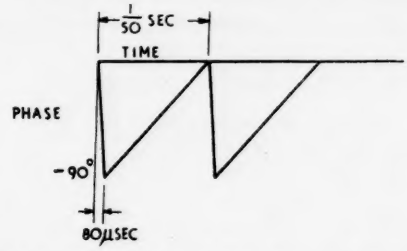


Fig. 3. The waveform of the phase modulation.

of the spikes leads the main doppler trace, the echo is approaching, and if it lags the main doppler trace, the echo is receding. The spikes also give $\frac{1}{50}$ sec. time markers, by means of which the doppler frequency can be measured.

This technique is basically the same as that of MANNING, VILLARD and PETERSON (1950) who used two receivers rather than a phase modulated system.

1.3 *Range measurement*

Pulses are used to give the range of the echoes. The final amplifier valves of the transmitter are amplitude modulated with 10μ sec. pulses at a repetition rate of 100 pps. During the pulse, the output power of the transmitter is raised from 250 watts to approximately 4 kilowatts. The ground pulse triggers a time base in the range display unit at Salisbury and the time base is applied to the vertical plates of a cathode ray tube, which is also intensity modulated with 20 kilometre range markers and with the output of the wide-band range receiver. The cathode ray tube is photographed on horizontally moving bromide paper, consequently the range markers and the echo appear as dotted horizontal lines. The distance, as measured in this way is clearly

$$\tfrac{1}{2}(AP + PS - AS),$$

where

AP = distance from Adelaide to the reflection point

PS = distance from reflection point to Salisbury

AS = distance from Adelaide to Salisbury.

The slant range from Salisbury can only be calculated with a knowledge of the direction of P. When the direction of arrival of the reflected wave is known, the calculation of the slant range is not difficult, and is given in § 1.6.

The range pulses do not influence the output of the doppler receiver. because of its limited bandwidth which is about 2 kc. Although some interference was anticipated, in practice no interference whatever was experienced from the pulses. A bandwidth as large as 2 kc is necessary in order to preserve the sense spikes. The bandwidth of the range receiver is 100 kc.

1.4 *Direction of arrival*

The direction of arrival of the reflected wave is measured by a phase comparison method. The method differs from the well known method of Ross. Bramley and Ashwell (1951) in that the phases of doppler signals received at spaced aerials are compared, rather than the radio frequency phases of the downcoming waves.

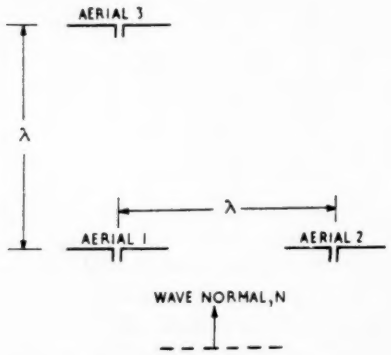


Fig. 4. Arrangement of the direction finding aerials.

Fig. 5. Vector diagram for illustrating the operation of the direction finding equipment.

Fortunately the radiofrequency phase relationships are preserved in the doppler beats so that no precautions are necessary to ensure equality of phase shift in the various receivers. Three identical dipole aerials are used, as shown in Fig. 4. Let N be the wave normal of the ground wave, which in practice is a plane wave.

The electric field of the ground wave is therefore in phase at the three aerials, and, neglecting coupling,* the currents induced in the aerials by the ground wave, will also be in phase.

Let the vectors in Fig. 5 be defined as follows:

G = the current induced in aerials 1 and 2 by the ground wave.

S_1 = the current induced in aerial 1 by the reflected wave.

S_2 = the current induced in aerial 2 by the reflected wave.

θ = the phase angle between S_1 and S_2.

R_1 = the resultant current in aerial 1.

R_2 = the resultant current in aerial 2.

In general, the reflected wave is shifted in frequency because of the motion of the trail. Suppose the point of reflection to be approaching. Then S_1 and S_2 will

* The effects of coupling have been calculated and found to be small enough to neglect in this experiment. The system has also been calibrated by means of an aeroplane and the calibration is given in § 4.

VOL. 4 1953

rotate in a counterclockwise direction relative to G. R_1 has passed its maximum value and R_2 is approaching its maximum value. The phase of the doppler signal received at aerial 1 will therefore lead the phase of the doppler signal received at aerial 2. Moreover, the maxima of the two doppler signals on the record will clearly be separated by the phase angle θ. The phase difference between the reflected signals incident at the two aerials is therefore the same as the phase difference between the corresponding doppler beats, for approaching echoes. The same reasoning applies to aerials 1 and 3.

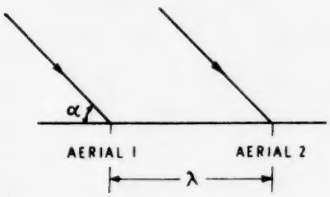


Fig. 6. Diagram showing the method of determining the direction of arrival, from the direction finding records.

For receding echoes, S_1 and S_2 rotate in a clockwise sense and R_2 leads R_1 in phase. Thus for receding echoes, if S_1 leads S_2 in phase by an angle θ, R_1 will lag R_2 by the same angle θ. The phase difference between the reflected signals incident at the two aerials is therefore equal and opposite to the phase difference between the corresponding doppler beats for receding echoes.

In Fig. 6. α is the angle between the direction of arrival of the reflected signal and the line joining aerials 1 and 2. The phase of the reflected wave incident at aerial 1 leads that of the reflected wave incident at aerial 2 by an angle $\theta = \dfrac{\lambda \cos \alpha}{\lambda} 2\pi$ radians.

$$\therefore \quad \theta = 2\pi \cos \alpha \quad \text{or} \quad \cos \alpha = \frac{\theta}{2\pi} \tag{2}$$

For an approaching trail θ is the angle by which the phase of the doppler beat received at aerial 1 leads that received at aerial 2.

For a receding echo θ is the angle by which the phase of the doppler beat received at aerial 1 lags that received at aerial 2.

With a knowledge of doppler sense, as determined from the spikes, and the angle θ, the space angle α is readily found. A second angle β, the angle between the direction of arrival of the reflected wave and the line through aerials 1 and 3 is found in the same way, and

$$\cos \beta = \frac{\theta'}{2\pi} \tag{3}$$

where θ' is the phase difference in radians between the doppler signals received on aerials 1 and 3.

1.5 *Wind measurement*

Choose the set of axes shown in Fig. 7, where the origin is the centre of aerial 1, the Z axis is a vertical line through the origin, the X axis is a line through aerial 2, and the Y axis is a line through the centre of aerial 3.

The direction cosines of the line OR which points in the direction of arrival of the echo are $\cos \alpha_1$, $\cos \beta_1$ and $\sqrt{1 - \cos^2 \alpha_1 - \cos^2 \beta_1}$. Cos α_1 and cos β_1 can be found directly from the relative phases of the doppler beats, with a knowledge of doppler sense. The azimuth of the point of reflection is given by

$$\phi_1 = \tan^{-1} \frac{\cos \beta_1}{\cos \alpha_1} \tag{4}$$

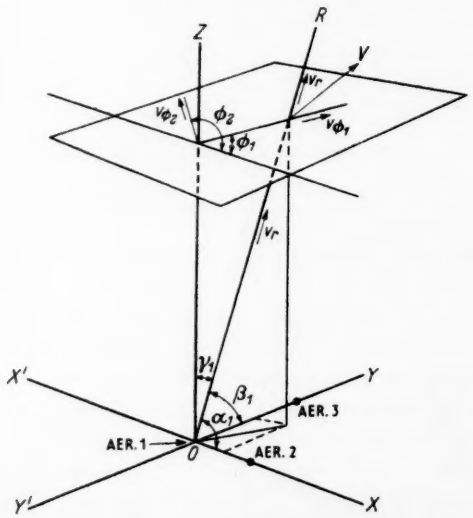


Fig. 7. The geometry of the direction finding system.

where the X axis is the reference line and the azimuth of OY is $+\pi/2$.

The zenith angle of the point of reflection is γ_1, where

$$\gamma_1 = \cos^{-1} \sqrt{1 - \cos^2 \alpha_1 - \cos^2 \beta_1} \tag{5}$$

It has been shown by Manning, Villard and Peterson, and confirmed by us, that the winds at this level are substantially horizontal. This simplifies the problem of wind measurement considerably. If the wind is assumed to be horizontal, the occurrence of two meteors within a short interval of time and at approximately the same height provides enough information to define the wind vector completely at that height and time.

The angles α_1 and β_1 (Fig. 7) can be found directly by phase comparison of the doppler beats received at the three aerials, with a knowledge of doppler sense. Thus the azimuth ϕ_1 and zenith angle γ_1 of the line OR can be found. Let a receding echo occur in the direction OR with a line of sight component of velocity v_r. If the wind vector V is horizontal, its component in direction ϕ_1 is given by $v_{\phi_1} = \dfrac{v_r}{\sin \gamma_1}$.

Thus, a single meteor gives the component of wind velocity in the direction of its point of reflection. A second meteor will, in general, give the component of V in another direction ϕ_2. Two such components are shown in Fig. 8 and the horizontal wind with these components is clearly given by V.

VOL. 4 1953

Usually, meteor echoes do not occur close enough in time and height to follow the wind continuously in this way. Occasionally the echo rate is large enough for this method of analysis to be applicable, but the azimuth angles of many of the pairs do not differ sufficiently to give reasonable accuracy. This is likely to happen during showers, unless there are a number of radiants active at the same time.

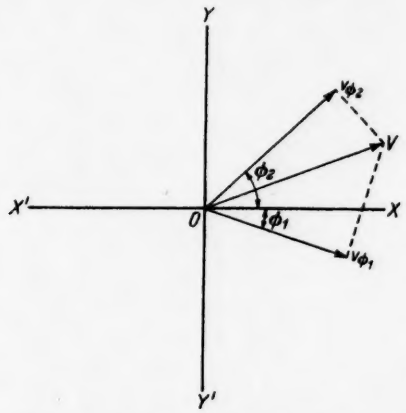


Fig. 8. The method of finding the wind vector from two meteor echoes.

The average diurnal behaviour of the wind at different times of the year is of great interest, and the average winds are found as follows. The apparatus is allowed to run for several weeks, and the results are grouped into 24 hourly intervals, each hourly interval containing information from all the days. For each hourly interval the components v_ϕ are grouped and the wind vector V which best fits this set of components is calculated and in this way all the recorded echoes are used. If the heights of the echoes are known, the information can be split into height groups, so that variations in the average diurnal winds with height can be studied. The method of obtaining the best fitting vector V is essentially that of MANNING, VILLARD and PETERSON (1950), as outlined in paper II.

1.6 *Height determination*

The height of the point of reflection is found as follows. In Fig. 9, A is the transmitter, S the receiver and P is the reflection point. The angle β is found directly by phase comparison of the signals received on aerials 1 and 2. The distance measured on the range record is given by

$$r' = \tfrac{1}{2}(AP + r - AS).$$

AS is small compared with r, so that $AP \doteq r + AS\cos\beta$.

$$\therefore \quad r' \doteq \tfrac{1}{2}[2r - AS(1 - \cos\beta)]$$

$$\text{or} \quad r \doteq r' + \tfrac{1}{2}AS(1 - \cos\beta) = r' + 9{\cdot}8(1 - \cos\beta) \text{ kilometres} \qquad (6)$$

where r' is the trace range in kilometres.

The height of the echo is given by $h = r\cos\gamma$ kilometres.

1.7 *Simplified method*

Analysis of a large number of records for the summer months indicates that the most important features of the wind system in the meteor region are independent of height (*vide* Paper II). If the measurement of height is not required, the

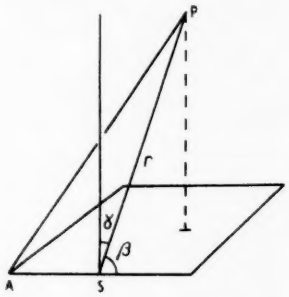


Fig. 9. Diagram to show the method of determining height.

apparatus can be greatly simplified, as the direction finding equipment then proves sufficient without sense or range determination. The equipment then reduces to a simple C.W. transmitter and a number of narrow band receivers with appropriately spaced aerials.

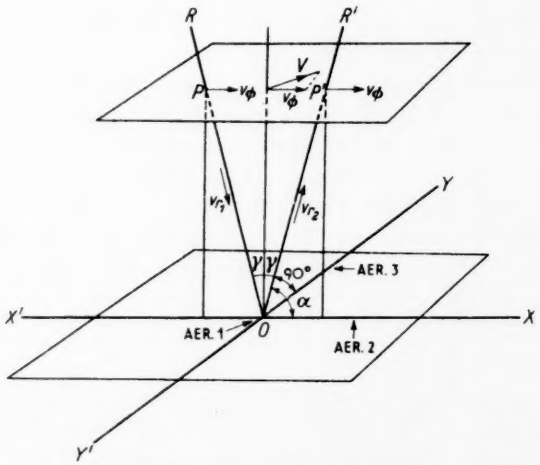


Fig. 10. Diagram illustrating the simplified method for wind determination.

Consider the operation of the direction finding equipment without the aid of the sense spikes. It is then not possible to say whether an echo is receding or approaching, and an ambiguity arises in the direction finding records.

Consider the simplified case in Fig. 10.

Let an echo occur at any point P on the line OR whose azimuth lies along OX' and assume the existence of a horizontal wind V. With this arrangement, the echo will be approaching and therefore the phase of the doppler beat received at aerial 1 will lead that received at aerial 2 by an angle $\theta = 2\pi \cos \alpha$. However,

VOL. 4 1953

an echo occurring at any point P' or OR' whose azimuth lies along OX will be receding, so that the phase of the doppler beat at aerial 1 will again lead that received at aerial 2 by the same angle $\theta = 2\pi \cos \alpha$. It is therefore impossible to distinguish between an approaching echo at P and a receding echo at P' without a knowledge of doppler sense.

Suppose the phase of the doppler beat received at aerial 1 leads that received at aerial 2 by an angle $\theta = 2\pi \cos \alpha$. Let the doppler beats received at aerials 1 and 3 be in phase with one another, so that $\beta = \frac{\pi}{2}$. Without the sense spikes, we can only say that the reflection point is either (1) on the line OR and approaching or (2) on the line OR' and receding.

The line of sight component of velocity v_r is known from the frequency of the doppler beat. In both cases, the horizontal component of V along the X axis will be $v_\phi = \frac{v_r}{\sin \gamma}$ and will be in the direction $+OX$.

Thus, the wind information is not affected by this ambiguity in the direction finding equipment, and it is of no consequence in wind measurement whether the echo is an approaching echo in one direction or a receding echo in the reciprocal direction. The direction of the echo is, of course, important for height determination.

1.8 *Ambiguity in D.F. records*

Unfortunately, another ambiguity arises in the three aerial direction finding system. When the phase difference between a pair of doppler beats approaches π, it is not possible to say with any certainty which signal leads which, in phase. The doppler beats received at aerials 1 and 3 differ in phase by π, when $2\pi \cos \alpha = \pi$, or $\alpha = 60°$. One solution of this difficulty would be to beam the transmitted power into a narrower cone of radiation. Another solution is to use an additional pair of doppler aerials with smaller spacing, and this has been done.

A diagram of the complete direction finding aerial system is shown in Fig. 11. The centres of aerials 4 and 5 are on a line parallel to the line joining aerials 2 and 3. The ground wave at aerial 4 leads that at aerial 5 by π and not 2π. Therefore echoes occurring on the plane S will produce out of phase doppler beats at aerials 4 and 5. The doppler beats are applied to the vertical plates of a double beam oscilloscope in which an out of phase signal causes an in phase deflection. Therefore echoes occurring on the plane S produce in phase doppler beats on the records.

Approaching echoes which occur on the side D_1 of S produce doppler signals at aerial 5 which lead those at aerial 4. Approaching echoes on the side D_2 of S produce doppler signals at aerial 5 which lag those at aerial 4. As before, the situation is reversed for receding echoes. It is thus possible to tell on which side of the plane S any point of reflection lies. To resolve completely all ambiguities, a sixth aerial would be required, but the additional information supplied by aerials 4 and 5 is sufficient to resolve the majority of ambiguities that arise. The operation of aerials 4 and 5 is affected considerably by their mutual coupling, however they are used in a purely qualitative manner on echoes with large zenith angles, and the

coupling is not serious enough to prevent them from being used in this manner.

The direction finding equipment has been calibrated by means of an aeroplane, and details of its calibration are given in § 4.

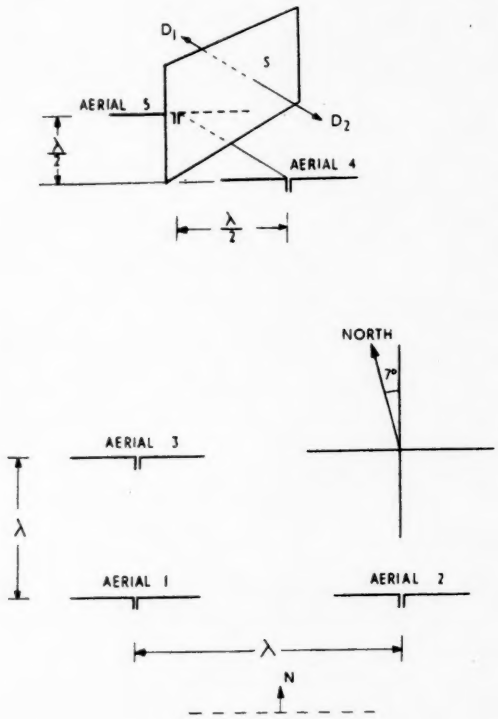


Fig. 11. Diagram of the complete direction finding aerial system.

2. Experimental Equipment

2.1 *The transmitter*

The transmitter is crystal controlled, the crystal frequency being 743·7 kc. The output of the crystal oscillator is phase modulated with the waveform of Fig. 3, the deviation being $2\frac{1}{2}°$. The frequency is then multiplied by 36, giving a frequency of 26·77 mc with a phase deviation of 90°.

The power amplifier has a continuous wave output of 250 watts, and is amplitude modulated with 4 kilowatt 10μsec pulses at a repetition rate of 100 pps. A pair of VT90 valves is used in the power amplifier, these valves having enough emission and electrode spacing to handle the pulses and enough grid and plate dissipation to run continuously. The transmitting aerial consists of two dipoles fed in phase. The dipoles are mounted in a horizontal plane and have parasitic reflectors underneath them.

VOL. 4 1953

2.2 *The receivers and recorder*

A block diagram of the receiving and recording equipment is given in Fig. 12. There are four narrow-band doppler receivers and one wide-band pulse receiver. The outputs of the doppler receivers are applied to the vertical plates of cathode ray tubes and photographed on horizontally moving bromide paper. An intensity modulated cathode ray tube is used for recording the range on the same paper.

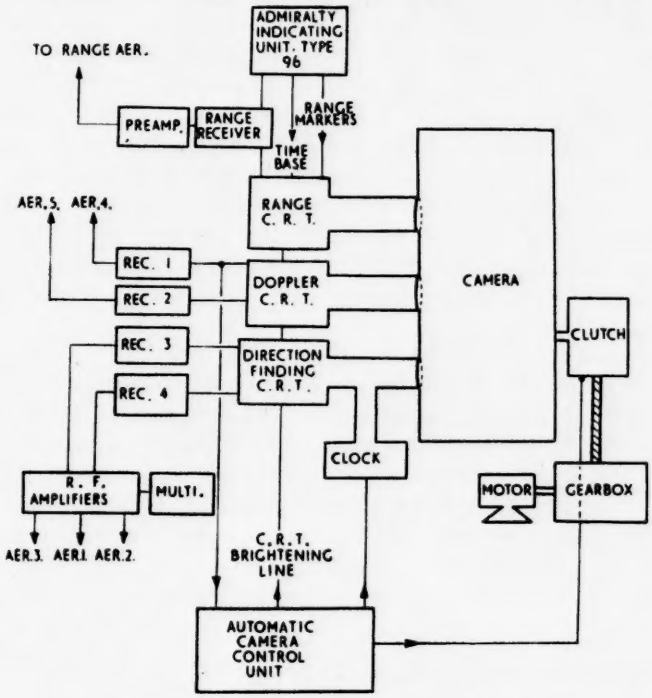


Fig. 12. Block diagram of the equipment at Salisbury.

The recording camera has three lenses, one for each cathode ray tube. There are two double beam tubes for the four doppler receivers and a single beam tube for the range receiver. The photographic paper is 7 cm wide, and each lens uses a third of the paper. The paper travels at 3.5 cm/sec.

Typical records are shown in Fig. 13. The range trace appears at the top, the doppler signals from aerials 4 and 5 are in the middle and the main direction finding traces are at the bottom of the paper. The horizontal distance between lenses is ten centimetres, so that the three images are laterally separated by this distance. For the purpose of publication, the traces shown in Fig. 13 have been aligned vertically.

Meteor echoes usually start suddenly and die away gradually and the recording cycle is started electronically by any sudden change in the output of doppler receiver I. Relays with various time delays then start the following sequence of events.

1. The paper starts moving and the cathode ray tubes brighten.

2. The paper continues to move until one second after the signal from receiver 1 drops below a predetermined level.

3. Then the paper stops, the cathode ray tubes are blacked out and lights are flashed on the clock face.

4. The paper moves on for half a second, and waits for the next echo.

The paper is started and stopped by means of an iron dust and oil magnetic clutch, operated by five power valves in parallel, which produce a current of half an ampere through the windings. The recording cycle is completely automatic so that the equipment is unattended and runs continuously.

Only two receivers are used in conjunction with the three direction finding aerials. Aerials 1 and 3 are switched in turn to receiver 3, while aerials 1 and 2 are switched in turn to receiver 4. The switching is done by means of a square wave applied to the grids of the preamplifiers. The square wave is asymmetrical, so that one aerial is connected to its receiver for a longer period than the other; thus producing a dark trace and a light trace on the recording paper for each pair of aerials. The trace appropriate to each aerial of a pair can then be identified.

3. Interpretation of the Records

The line through aerials 1 and 3 of Fig. 11 points in the direction 7° East of North, while the line through aerials 1 and 2 points in a direction 7° South of East. For convenience, the pairs of aerials will therefore be referred to as the North-South and East-West pairs respectively, aerial 1 being common to both pairs.

VOL. 4 1953

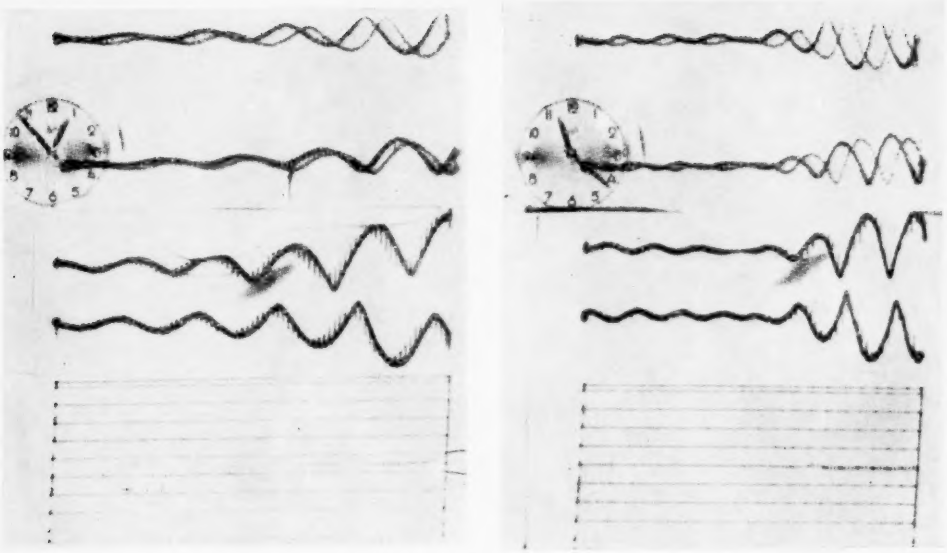
It is desirable to calculate the results in terms of the vector direction of the wind, and for this reason, receding echoes will be regarded as having positive signs, although they show negative doppler shifts.

Referring to Fig. 13a, the traces on the record are arranged as follows:

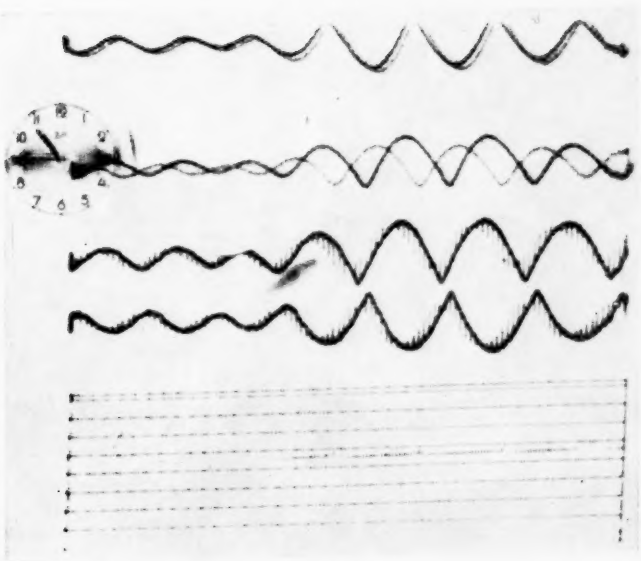
Trace 1	Range
Trace 2	Aerial 5 (North West aerial)
Trace 3	Aerial 4 (South East aerial)
Trace 4	North-South pair Dark trace—Aerial 3 or North aerial Light trace—Aerial 1 or South aerial.
Trace 5	East-West pair. Dark trace—Aerial 1 or West aerial Light trace—Aerial 2 or East aerial.

Phase comparisons of the dark and light doppler signals on traces 4 and 5 therefore give $\cos\beta$ and $\cos\alpha$ respectively (Fig. 14). The azimuth angle ϕ. is given by $\phi = \tan^{-1}\dfrac{\cos\beta}{\cos\alpha}$, where the line through aerials 1 and 2 is the reference line and the line through aerials 1 and 3 has an azimuth of $+90°$.

The phase relationship between traces 2 and 3, i.e. between aerials 5 and 4, is used to resolve the uncertainty which arises when the phase difference between a dark and light trace approaches 180°.



a b



c

Fig. 13a, b, c. Some typical records.

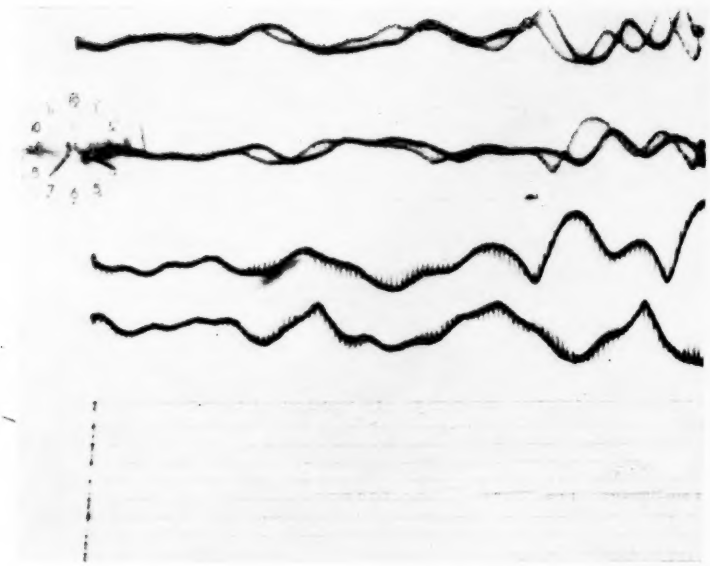


Fig. 13e. A typical record.

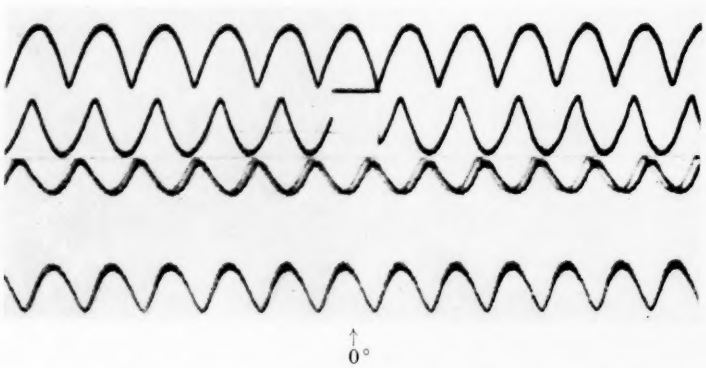


Fig. 15. Record from an aeroplane flying across the receiving site from North to South.

VOL.
4
1953

For receding echoes, the direction finding records are interpreted as follows:

Trace 4:

Light trace leads dark trace in phase—Northern echo $\therefore \quad \beta < 90°$.

Dark leads light—Southern echo $\therefore \quad \beta > 90°$.

Trace 5:

Dark leads light—Eastern echo, $\alpha < 90°$.

Light leads dark—Western echo, $\alpha > 90°$.

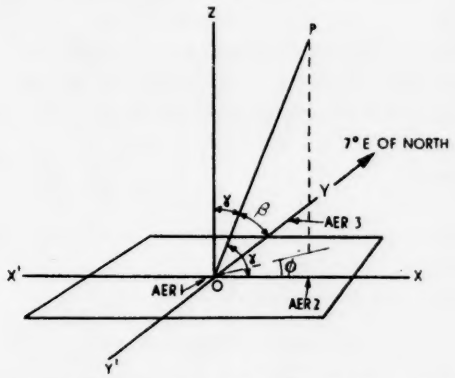


Fig. 14. Orientation of the direction finding system.

If trace 3 leads trace 2 in phase, the echo lies on the side D_1 of plane S in Fig. 12.

If trace 2 leads trace 3 in phase, the echo lies on the side D_2 of S.

It is of advantage to consider all echoes as receding when reading the direction finding traces. For approaching echoes, this gives the reciprocal bearing which is then reversed to give the true bearing.

The record of Fig. 13a, can now be analyzed. The magnitude of the doppler shift can be found from any of the doppler traces. The paper speed is 3.5 cm/sec and the wavelength of a doppler cycle is 1·2 cm. The doppler frequency is therefore 2·9 cycles/sec. and, from equation 1, this corresponds to a line of sight component of drift velocity of 16 metres/sec

$$\therefore \quad v_r = 16 \text{ metres/sec.}$$

On trace 4. light leads dark by 0·14 cycles and the echo is therefore a Northern echo. and from equation 3, $\cos\beta = +0{\cdot}14$.

On trace 5. light leads dark by 0·26 cycles, the echo is therefore a Western echo. and $\cos\alpha = -0{\cdot}26$.

In addition. trace 3 leads trace 2 in phase, and this is consistent with a North West echo.

From equation 4, the azimuth angle $\phi = \tan^{-1}\dfrac{\cos\beta}{\cos\alpha} = \tan^{-1}\dfrac{+0{\cdot}14}{-0{\cdot}26} = 152°$,

or 28° North of West.

From equation 5, $\cos\gamma = \sqrt{1 - \cos^2\alpha - \cos^2\beta} = 0{\cdot}95$

$$\therefore \quad \gamma = 18^\circ \quad \text{and} \quad \sin\gamma = 0{\cdot}31$$

$$\therefore \quad v_\phi = \frac{16}{0{\cdot}31} = 52 \text{ metres/sec}$$

Assuming a horizontal wind, the direction finding traces therefore tell us that the component of the wind in the direction 28° North of West was 52 metres/sec. The time on the record was 00·53, and it was taken on October 30th. This result is quite definite, in spite of the fact that the sense spikes and range trace have not yet been used.

The sense spikes lag the main doppler trace in phase, so that the echo is receding, and the above value of ϕ is the true azimuth and not its reciprocal. The range reading is approximately 84 kilometres. Therefore, from equation 6, the slant range r is given by

$$r = 84 + 9{\cdot}8(1 - 0{\cdot}14) = 92 \text{ kilometres.}$$

The height is therefore $92 \cos\gamma = 92 \times 0{\cdot}95 = 87$ kilometres.

Fig. 13b is analysed as follows:

The doppler frequency is 4·9 cycles/sec

$$\therefore \quad v_r = 27 \text{ metres/sec.}$$

On trace 4 light leads dark by 0.36 cycles. The echo is therefore a Northern echo and $\cos\beta = +0{\cdot}36$.

On trace 5, light leads dark by 0·40 cycles. The echo is therefore a Western echo and $\cos\alpha = -0{\cdot}40$.

Trace 3 leads trace 2 in phase and this is consistent with a North-West echo.

$$\phi = \tan^{-1}\frac{+0{\cdot}36}{-0{\cdot}40} = 42^\circ \text{ North of West}$$

$$\cos\gamma = \sqrt{1 - (0{\cdot}36)^2 - (0{\cdot}40)^2} = 0{\cdot}84 \quad \therefore \quad \gamma = 33^\circ \quad \text{and} \quad \sin\gamma = 0{\cdot}54$$

$$\therefore \quad v_\phi = \frac{27}{0{\cdot}54} = 50 \text{ metres/sec}$$

Assuming a horizontal wind, its component in the direction 42° North of West was therefore 50 metres/sec at 23·42 on October 29th.

The echo is approaching, therefore the above value of ϕ is its reciprocal bearing. The true azimuth of the echo is therefore 42° South of East. The trace range is 93 kilometres, the range is therefore given by

$$r = 93 + 9{\cdot}8(1 + 0{\cdot}36) = 106 \text{ kilometres.}$$

Its height is given by

$$h = 106 \cos\gamma = 89 \text{ kilometres.}$$

The echo in Fig. 13c occurred at 22·45 on October 29th. Dark leads light on

VOL. 4 1953

trace 4, by 0·38 cycles, and light leads dark on trace 5 by 0·07 cycles. One would therefore say, at first glance, that the azimuth of the echo was slightly West of South. However, the phase relationship between traces 2 and 3 show that this is incorrect. Trace 3 leads trace 2 in phase, so that the echo lies on side D_1 of S in Fig. 11, whereas Southern echoes clearly lie on side D_2 of S.

The correct interpretation of the record is therefore as follows:

On trace 4, light leads dark by 0·62 cycles. This gives a Northern echo and $\cos\beta = +0{\cdot}62$.

On trace 5, light leads dark by 0·07 cycles. This gives a Western echo and $\cos\alpha = -0{\cdot}07$.

A North West echo is consistent with the phase relationship between traces 2 and 3.

$$\phi = \tan^{-1}\frac{+0{\cdot}62}{-0{\cdot}07} = 84^\circ \text{ North of West}$$

$$\cos\gamma = \sqrt{1 - (0{\cdot}62)^2 - (0{\cdot}07)^2} = 0{\cdot}78$$

$$\therefore \quad \gamma = 39^\circ \quad \text{and} \quad \sin\gamma = 0{\cdot}63$$

The doppler frequency is 2·9 cycles/sec, so that $v_r = 16$ metres/sec

$$\therefore v_\phi = \frac{16}{0{\cdot}63} = 26 \text{ metres/sec.}$$

The echo is receding, therefore the above value of ϕ is its true value and not its reciprocal. The trace range is 107 kilometres.

$$\therefore \quad r = 107 + 9{\cdot}8(1 - 0{\cdot}62) = 111 \text{ kilometres}$$

and

$$h = 111\cos\gamma = 87 \text{ kilometres.}$$

The analysis may appear to be somewhat lengthy, but, with practice the azimuths can be roughly estimated at a glance, the longest part of the analysis being the measurement of the phase differences. The arithmetic is simplified by the use of graphs from which $\frac{1}{\sin\gamma}$ and $\cos\gamma$ can be read directly with a knowledge of $\cos\beta$ and $\cos\alpha$. In addition to this, a table is used to give values of the quantity $9{\cdot}8(1 - \cos\beta)$ for various values of $\cos\beta$. With such aids, and with practice, the interpretation of records becomes a routine procedure.

A confused echo is shown in Fig. 13d, and such are rejected for wind analysis. These echoes are usually long enduring and apparently involve scattering from more than one centre.

4. Calibration of the direction finder

The axes of the direction finding system are in the direction 7° East of North and 7° South of East. An aeroplane was flown on these bearings across the receiving site at a height of 4,000 feet. Several runs were made, and the best runs were used for the calibration.

A part of the record which was taken during the best South to North run is shown in Fig. 15. The position of the aircraft was checked optically, the markers on trace 3 being made by the observer at intervals of five degrees in the zenith angle of the aircraft.

Since the transmitter is at Adelaide, which is South of Salisbury, an aircraft flying from South to North does not give a zero doppler frequency when passing over the receiving site. When overhead and flying North, the distance from the transmitter to the aircraft is increasing, and the distance from the aircraft to the receivers is not changing. Therefore, the total distance is increasing with time, or the aircraft is apparently receding.

The part of the record shown in Fig. 15, was taken when the aircraft was passing overhead. The East-West pair of traces are in phase, but on the North-South trace, light leads dark by 0·1 cycles, giving $\cos\beta = +0{\cdot}1$. This corresponds to an error of about 6° in zenith angle.

If 0·1 is subtracted from the value of $\cos\beta$, the error is removed for overhead echoes. It is also reduced considerably everywhere else, for after making this correction, the error is less than 4° for all zenith angles less than 40°. The average error in the corrected zenith angle is 1·2° over the South to North run and 2° over the West to East run.

Acknowledgements—The authors wish to thank Professor L. G. H. Huxley for suggesting the field of work and for his constant encouragement and advice, and also Mr. A. A. Weiss for his assistance in the early stages of the work. They are indebted to Mr. B. A. Chartres for designing the camera and to Messrs. J. Gascoigne and S. O. Sincock for their part in building the clutch, gearbox and camera.

The authors also wish to thank the Department of Supply for providing the receiving site at Salisbury, South Australia, and for assistance with equipment, and the Electrical Engineering Department of the University for the loan of a transmitter.

One of us (D.S.R.) is indebted to the *Imperial Chemical Industries* (Aust. and N.Z.) for the award of a Research Fellowship.

This work was financed by the University of Adelaide and the Radio Research Board of Australia.

VOL.
4
1953

References

Greenhow, J. S.	1952	J. Atmosph. Terr. Phys. **2**, 282
Manning, L. A., Villard, O. G. and Peterson, A. M.	1950	Proc. Inst. Radio Engrs. **38**, 887
Ross, W., Bramley, E. N. and Ashwell, G. E.	1951	Proc. Inst. Elect. Eng. Part 3 **98**, 294

Journal of Atmospheric and Terrestrial Physics, 1953, Vol. 4, pp. 271 to 284. Pergamon Press Ltd., London

Measurements of winds in the upper atmosphere by means of drifting meteor trails II

W. G. Elford and D. S. Robertson

Department of Physics, The University of Adelaide, South Australia.

(*Received* 10 *June* 1953)

ABSTRACT

Winds in the upper atmosphere have been deduced from the drifts of meteor trails by means of a continuously recording doppler-radar system operating on 27 mc. Measurements of the average diurnal variations of the winds at known heights between 80–105 km have been obtained for each of the three months, October, November and December, 1952. The winds are found to be essentially horizontal, and analysis of the diurnal variations shows the presence of prevailing and periodic components of the wind. The 12 hr and 24 hr harmonic components represent anticlockwise rotation of the wind vectors; the 12 hr component is consistent with the phase of a semi-diurnal tidal wind as deduced from barometric oscillations. While, at any time, the direction of the wind remains the same over the height range investigated, the mean velocity of the wind increases with height.

1. Introduction

A number of investigations of winds in the upper atmosphere using radio techniques have been made in recent years, but these measurements have been, in the main, confined to the *E* and *F* regions of the ionosphere since they depend upon the presence of irregularities in these regions (Beynon, 1948; Munro, 1950; Phillips, 1952; Salzberg and Greenstone, 1951). A limited number of measurements have also been made on the movements of sporadic *E* ionization (Gerson, 1950; Ferrell, 1948).

In the height range 80–100 km the recording of radio echoes from meteor trails leads to reliable methods of measuring winds. Two techniques have been used. Manning, Villard and Peterson (1950) have described a C.W. doppler system on 23 mc and Greenhow (1952) has used a pulse, range-drift method on 36 mc.

In what follows, the results of an investigation of winds as deduced from the records of drifting meteor trails over a period of three months are presented and analysed. The observations of the meteor trails are made by means of a self-recording doppler-radar system which embodies some novel features and is described in a companion paper by Robertson, Liddy and Elford (henceforth referred to as Paper I).

The average diurnal wind variations for each of the three months have been obtained and the prevailing and harmonic components of the winds extracted. An analysis has also been made of the average vertical wind motion and the velocity gradient of the wind ir this region.

2. Techniques

Although a full description of the techniques and apparatus is given in Paper I, yet for convenience they are again summarized below.

To determine the motion of the upper atmosphere from the drift of meteor trails it is necessary to know the line of sight velocity of the trails relative to an observer on the ground. and the position of the trails in space. These parameters

are measured by a combined C.W. and pulse technique on a frequency of 27 mc. The transmitter is situated at Adelaide and radiates a C.W. signal of 250 watts within a vertical cone of 40°. The receiving site is at Salisbury (19·6 km north of Adelaide) where the doppler beat between the ground wave and the reflected wave from the meteor trail is automatically recorded. The line of sight velocity of drift is deduced directly from the doppler record. To determine the sense of the drift the phase of the transmitted wave is abruptly retarded 90° every $\frac{1}{50}$ of a second and the result is to superimpose spikes on the doppler record, the locus of whose ends is a waveform which either leads or lags the main doppler waveform according as the trail approaches or recedes. (§ 1.2, Paper I).

To determine the position of the trail in space, both the direction of arrival of the sky wave and the slant range of the reflection point must be measured. The latter is determined by pulse-modulating the transmitted signal above the C.W. background and recording the time delay between the ground wave and the sky wave on a conventional radar display. Measurement of the direction of arrival of the sky wave is made by using three receiving aerials at the corners of a right-angled triangle. These aerials are so positioned that the ground wave is in phase at all three aerials (i.e. the phase differences are 0, 0, and 360°) and the relative phases of the sky wave at each aerial are recorded as the relative phases of the doppler beats at the detectors of the receivers. The centres of the aerials define the axes OX and OY of a Cartesian system in which OZ is vertical. The relative phases of the doppler beats in pairs of aerials give the direction cosines, in this system, of the wave normal of the sky wave reflected from the meteor trail.

All recording is automatic, the information being displayed on three cathode-ray tubes and photographed on a roll of bromide paper. The first echo from the meteor trail initiates the movement of the paper which continues to move at the rate of 3·5 cm per sec until the echo disappears. The paper then comes to rest and a clock face is photographed on to it.

Under normal operating conditions one 100 ft length of paper lasts for about 24 hours and records 200–300 meteor echoes of which about 25% are usable for analysis.

3. Observations and Reduction of Recordings

The observations were made at Adelaide (139° E, 35° S) during October 13–30, November 12–30, and December 2–20, 1952. With the exception of three hours each afternoon during October and occasional hours when the recording paper was used up before the camera was reloaded, the observations were continuous between the above dates.

The incidence of various meteor showers during the three months and the diurnal variation of the rate of sporadic meteors cause the number of echoes per hour to vary markedly during each day and from day to day. In general the activity is a maximum during mid-morning and a minimum during mid-afternoon.

Since the meteor trails are highly aspect sensitive, the reflection points during a meteor shower are nearly all located in only one half of the sky. At any time the reflection points for shower meteors lie in a plane through the observer and at right angles to the meteor trails.

VOL.
4
1953

Radio echoes from meteor trails with a single scattering centre give a regular doppler signal as shown in Fig. 1, Paper I, and only echoes of this type were used in the analysis. At least two cycles of dopper signal are required for analysis and about 25% of the records meet this requirement.

The direction cosines, found as mentioned above, are corrected for a small site error as explained in Paper I. The zenith angle and height of the reflection point are calculated from these direction cosines and the slant angle.

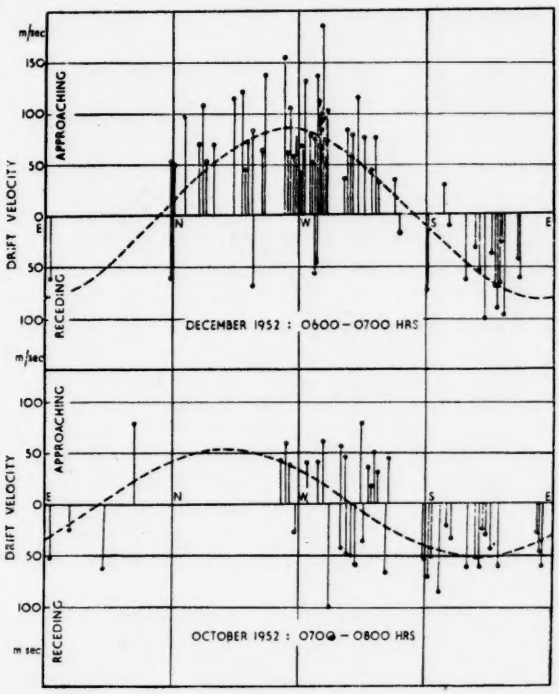


Fig. 1. Vector plots of the horizontal radial drift velocity versus azimuth for two hourly intervals. The dashed curves show the most probable value of the amplitude and direction of the wind.

At first it was assumed that the wind motion was horizontal, a justifiable assumption, since echoes from almost overhead showed little or no doppler shift. (The analysis of § 4.3 substantiates this assumption.) Hence from the line of sight velocity v_r and the zenith angle γ, the horizontal radial drift velocity v_ϕ is defined by $v_\phi = v_r/\sin\gamma$ and given the appropriate sign.

It is thus possible to measure the true wind given the position and radial drift of two meteor trails occuring close together in height and in time and at least 45° apart in azimuth. In general however the echo rate is too small to give the wind by this method, but in any case it is only of limited value. (*vide* § 4.2).

The chief feature of interest in any wind motion in this region is its average diurnal variation. If it is assumed that the wind on the average does not change rapidly in direction or amplitude from hour to hour, a simple averaging process can

be effected by grouping all the measurements obtained from one month's observation that occur in the same hourly interval of local time. In this way all the data for one month are collected into 24 hourly intervals, each interval containing measurements taken during that hour on as many as 18 consecutive days.

By plotting the horizontal radial drift velocity v_ϕ against the azimuth angle ϕ for each meteor echo within a given hourly interval, a vector plot is formed as shown in Fig. 1. Were these vectors to represent the radial components of a wind of fixed amplitude and direction, then their end points would lie on a sine curve, for the following reason.

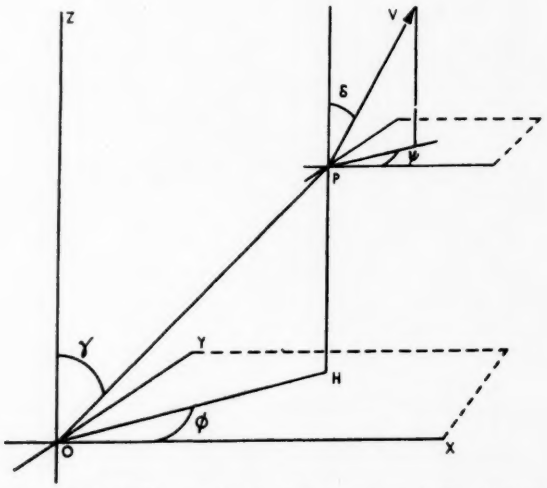


Fig. 2. Coordinate system for a drifting meteor trail. P represents the scattering centre of the trail and V the wind vector.

With reference to Fig. 2, the Cartesian coordinate system O, X, Y, Z is determined by the receiving aerials whose axes lie along OX and OY. The point P represents the scattering centre of a meteor trail which is drifting in the direction of the vector wind V.

The direction cosines of V are:

$$(\sin\delta\cos\psi,\ \sin\delta\sin\psi,\ \cos\delta)$$

and the direction cosines of OP are:

$$(\sin\gamma\cos\phi,\ \sin\gamma\sin\phi,\ \cos\gamma).$$

The component of V along OP is

$$v_r = V[\sin\delta\sin\gamma\cos(\phi-\psi) + \cos\delta\cos\gamma]$$

and the component of V along OH, whose direction cosines are $(\cos\phi, \sin\phi, 0)$, is

$$v_\phi = V\sin\delta\cos(\phi-\psi).$$

If V is horizontal, $\delta = 90°$, and the horizontal radial velocity becomes

$$v_\phi = V\cos(\phi-\psi).$$

VOL. 4 1953

Vector plots of the experimental values of v_ϕ against ϕ (Fig. 1) show general agreement with the above equation. The scatter observed can be attributed to several causes.

(i) Day to day variations in the amplitude of the wind and fluctuations in direction,

(ii) Variation of the mean velocity with height,

(iii) Random wind components, both horizontal and vertical.

Factors (i) and (iii) are inherent in the wind itself but the scatter due to (ii) could be reduced by grouping the data into height intervals. In the reduction of the observations the data were analysed in two height groups, 85–94 km, and 95–104 km, and also analysed as a whole.

To determine the most probable direction ψ and amplitude V of the wind for each hourly interval, a best fitting cosine curve was determined according to the "least square analysis" used by MANNING, VILLARD and PETERSON (1950) Using the notation of Fig. 2, if the equation of the curve to be fitted to the experimental ordinates is $v_\phi = V\cos(\phi - \psi)$, then

$$\tan\psi = \frac{\Sigma c^2 \Sigma v_{\phi_n} s - \Sigma cs \Sigma v_{\phi_n} c}{\Sigma s^2 \Sigma v_{\phi_n} c - \Sigma sc \Sigma v_{\phi_n} s}$$

and

$$V = \frac{\Sigma v_{\phi_n} c(1 + \tan^2\psi)^{\frac{1}{2}}}{\Sigma c^2 + \tan\psi\, \Sigma sc}$$

where v_{ϕ_n} is the experimental ordinate in the direction ϕ_n, and $s = \sin\phi_n$, $c = \cos\phi_n$.

In Fig. 1 the best fitting cosine curves, found using the above expressions, have been superimposed on the vector plots. The upper diagram is representative of many hourly intervals during November and December, the scatter being chiefly due to amplitude variations. The lower diagram was chosen to represent one of the most "confused" hourly intervals and it indicates considerable fluctuation in the direction of the wind.

The uneven distribution of the ordinates in azimuth is common to all the hourly intervals, and is due to the large proportion of the echoes that, in these months, come from shower meteors with reflection points lying in a limited part of the sky as mentioned above. In general each hourly interval has information in about three quadrants.

The number of echoes per hourly interval varies throughout the three months from 6 to 70. Since it was considered that at least ten echoes were needed to yield reliable values of V and ψ, no intervals with less than ten echoes were analysed.

The direction finding system does not lie exactly along the N-S and E-W lines and a correction of $-7°$ has to be made to all the computed values of ψ.

4. RESULTS

4.1. *Average monthly diurnal wind variations*

Having calculated the amplitude and direction of the winds for each hourly interval, the wind vectors can be resolved along the E-W and N-S directions. The components for the total height groups for each of the three months are given in Fig. 3, where the vectors have been drawn to represent the vector direction of the wind. The

patterns of the winds for November and December are similar, but unlike that for October which shows little of the prevailing East vector wind present in the other months. The average prevailing vector winds for these months are shown in Fig. 4.

The first and second harmonics of the monthly components have been extracted,

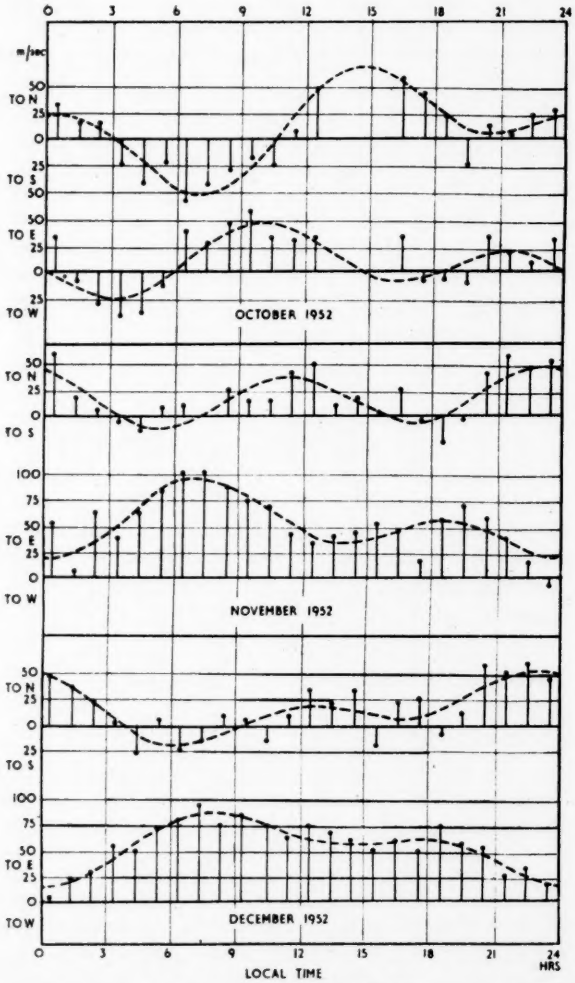


Fig. 3. Average diurnal variations of the N-S and E-W components of the winds.

and by adding these to the prevailing wind the dashed curves in Fig. 3 are obtained. The fit of these curves to the experimental ordinates indicates that the first and second harmonics are the only harmonics of any importance, the residuals being randomly distributed. Although it might be thought that the residuals would be largest for those hourly intervals containing the smaller numbers of meteor echoes, this is not so. In general the greatest scatter occurs for those intervals where the wind direction is changing most rapidly.

VOL. 4 1953

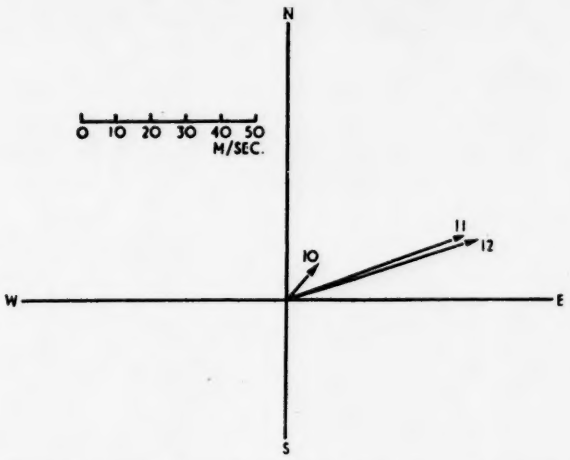


Fig. 4. Prevailing wind vectors for October, November and December, 1952.

During October no observations were made between 1300 and 1600 hours, consequently a Fourier analysis could not be made. However, the dashed curves for the N-S and E-W components are those compounded of a prevailing wind plus the first and second harmonics, which gave the best fit to the experimental ordinates.

In Fig. 5 the 12 and 24 hour harmonics are shown plotted on an harmonic dial.

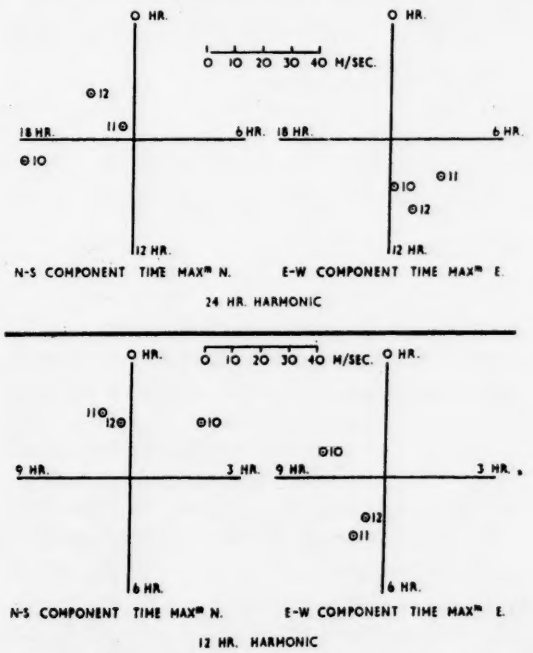


Fig. 5. Harmonic dials showing the 24 hr and 12 hr periodicity in the winds during October, November and December, 1952.

Although it is not justifiable to draw conclusions valid for a whole year from only three months of observation, it is of interest to note that,

(i) the wind vector associated with both harmonics rotates anticlockwise,

(ii) the diurnal wind component blows towards the North at 18 $\pm$ 2 hours,

(iii) the semi-diurnal wind component blows towards the North at 0 $\pm$ 1½ hours.

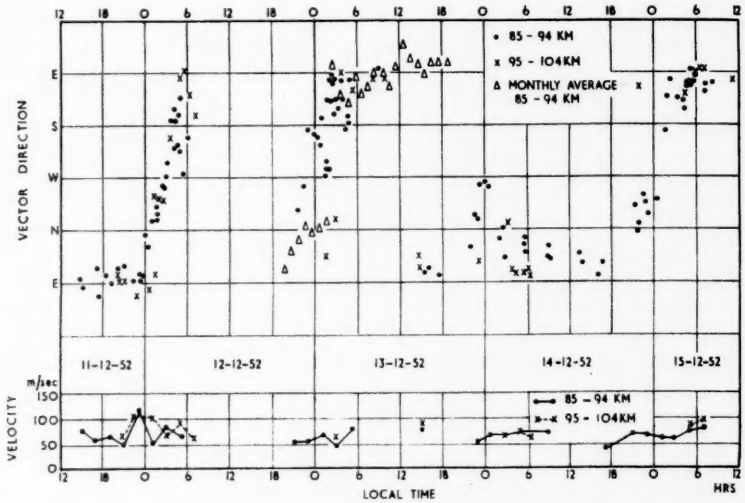


Fig. 6. Plot of the direction and amplitude of the wind for four consecutive days during December, 1952.

It can be seen from Fig. 3 that during November and December the phases of the harmonics were such that the two components reinforced the prevailing wind in the mid-morning causing the average velocity of the wind to increase to 100 m/sec towards the East. During October the maximum average wind velocity did not exceed 70 m/sec.

4.2. *Daily wind variations*

It is of interest to enquire how the wind varies from day to day and to compare it with the monthly average winds considered in the preceding section. The meteor rate is insufficient to permit this to be done in general, but during the Geminid shower in December the rate was adequate to allow a quasi-continuous recording of the wind for several days. Values of the radial components of velocity v_ϕ were determined from the echoes from two meteor trails which occurred in the same height group and less than one hour apart in time. The wind vector of which these two were components was then simply found from the point of intersection of the normals to these components drawn at their ends. Using pairs of meteor echoes in this way, the vector directions and average amplitudes of the winds were found for the two height groups, and the results are plotted in Fig. 6. The average azimuths derived from the hourly intervals for the whole month are also plotted for comparison. The Geminid shower culminated at approximately 0020 hours and con-

VOL. 4 1953

tinuous measurements could only be made for a few hours either side of culmination. However, these sufficed to indicate the main features of the wind.

During daylight the direction of the vector wind was mainly E to NE. which agrees well with the monthly averages. Between 2000 hours and 0600 hours a complete rotation occurred, this behaviour being well illustrated on the night of December 11–12. This rotation also took place on December 12–13 and 14–15, but it did not always occur at the same time. On December 12 the vector wind blew West at 0300 hours while on December 13 it blew West at midnight. This fluctuation in the time of rotation is reflected in the azimuths for the monthly hourly intervals, which indicate that the wind suddenly changed direction between 0200 and 0300 hours. From the Fourier analysis of the hourly vectors, given in the preceding section, it can be seen that this rapid change in direction is a rotation and is associated with the semi-diurnal component of the wind. It is apparent from the plot of December 13–14 that this rotation is not always completed, since on this night the wind began to rotate from East through North to West and then swung back again to the NE.

The azimuths for the two height groups show no significant difference, the wind direction behaving similarly from at least 85 to 105 km.

The manner in which the daily wind amplitudes vary is indicated by the plot of the amplitudes averaged over two hours. There is evidence of a sharp increase in the wind speed at approximately 2300 hours on December 11. These sudden increases in the speed of the wind are not uncommon and sometimes last for several hours.

4.3. *Vertical winds*

It has been mentioned in § 3 that meteor echoes from almost overhead showed little or no doppler shift and thus indicated the absence of vertical drift. However, the number of trails which occur within say 15°of the zenith is relatively small because the proportion of the total sky concerned is small. It is therefore desirable that the existence or absence of vertical drift should be more thoroughly examined.

In § 3 the line of sight velocity of a reflection point whose azimuth is ϕ and zenith angle γ is given by

$$v_r = V[\sin\delta \sin\gamma \cos(\phi - \psi) + \cos\delta \cos\gamma]$$

where ψ is the azimuth and δ is the zenith angle of the wind vector V.

If δ is not 90°, then the "apparent horizontal radial drift velocity" v_ϕ' is

$$v_\phi' = \frac{v_r}{\sin\gamma} = V[\sin\delta \cos(\phi - \psi) + \cos\delta \cot\gamma].$$

As before v_ϕ' varies sinusoidally with the azimuth angle ϕ, but in addition it contains a non-periodic term $V \cos\delta \cot\gamma$. Hence if the wind vector is not horizontal, the equation to be fitted to the experimental ordinates must be of the form.

$$v_\phi' = V_0' + V' \cos(\phi - \psi)$$

where

$$V_0' = V \cos\delta \cot\gamma$$

$$V' = V \sin\delta.$$

Let the angle of tilt of the wind vector V from the horizontal be τ.

Then $$\tau = 90 - \delta$$

Hence the angle of tilt is given by

$$\tan \tau = \cot \delta = \frac{V_0'}{V'} \tan \gamma$$

and the vertical wind component equals

$$V \cos \delta = V_0' \tan \gamma.$$

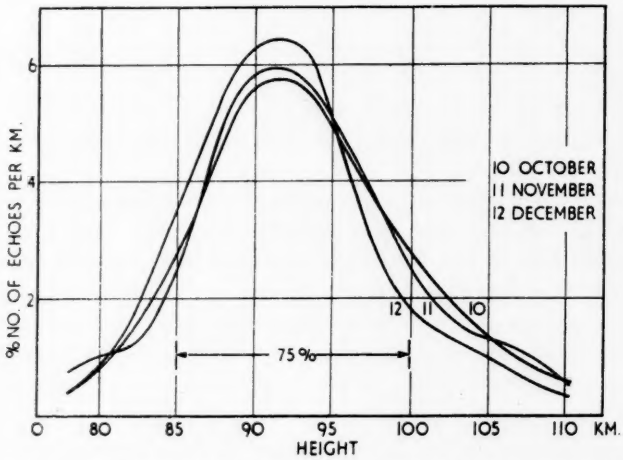


Fig. 7. Smoothed height distribution curves of the meteor echoes.

The equation for v_ϕ' has been fitted to several hourly intervals by the method of "least squares," and the value of $V \cos \delta$ determined, assuming an average value of 0·5 for $\tan \gamma$. The vertical wind component varied between $\pm$ 6 m/sec and showed no periodicity. The average value of the angle of tilt was $< 5°$.

It is concluded therefore, that the winds on the average are horizontal to within at least 5°. Since the above is an averaging process this result does not exclude the existence of random vertical components in the motion.

4.4. *Mean height of reflecting centres*

All echoes used in the analysis were distributed in height between 75 and 115 km, with 75% of them lying between 85 and 100 km. To test for variations of the winds with height, the data for each month were divided into two height groups, 85–94 km. and 95–104 km. The echoes were not distributed evenly throughout these height ranges, nor were the distributions in each hourly interval the same. Therefore, average heights for each hourly interval were first found and then the mean of these average heights was found for the total interval of 24 hours. In this way mean heights were found for the two height groups and also for the data as a whole. The results for the three months are given in Table 1. The probable error for all the heights in the table is $\pm$ 2 km.

VOL. 4 1953

The smoothed height distribution curves for the three months are shown in Fig. 7. They reveal that essentially the same region was investigated each month.

Table 1

Height Range / *Month*	85–94 km	95–104 km	75–115 km
October	90 km	98 km	93 km
November	90 ,,	98 ,,	93 ,,
December	90 ,,	98 ,,	92 ,,

4.5. *Changes of the wind speed with height*

The analysis of the wind data in two height groups has shown that the wind direction is essentially the same from 85 to 105 km. This fact is illustrated in Figs. 8a and 8b where the hourly vectors associated with the total group are compared with those for the two selected height groups for November and December. Although the data were in general insufficient to give a complete vector picture during the whole day for the two height groups, the vector components show similar trends. It was possible to make a Fourier analysis of the 85 to 94 km group for December and the resultant dashed curve closely resembles that for the total group.

A comparison of amplitudes of the average hourly vectors suggests that the amplitudes in the upper height group are greater than those in the lower group. In order to examine this suggestion, the ratios of amplitudes between the height groups for each hourly interval were formed, and the distributions of these ratios are plotted in Fig. 9a. Although the ratios show considerable scatter, their means are significantly greater than unity being larger, the greater the height separation. The average velocity of the wind for the total group was 61 m/sec for November and 64 m/sec for December, the mean heights from Table 1 being 93 and 92 km respectively. Using these values and the means of the amplitude ratios, the average velocities of the winds in the two selected height groups were calculated and are plotted in Fig. 9b. The results for both months show a positive wind gradient of 3·6 m/sec/km.

5. Discussion

5.1. *Prevailing winds*

Reference to Fig. 5 shows that the prevailing wind during November and December was directed towards the East while during October it was much smaller and directed towards the NE. This behaviour may well represent a seasonal change and can be compared with the properties of winds in the region 100–120 km found by the fading method. Phillips (1952) finds an average drift towards the East in summer and towards the West in winter.

5.2. *Periodic variations*

The analysis given in § 4.1 showed that both the 24 hour and the 12 hour harmonics represent anticlockwise rotation of the wind vector, such as would be anticipated for a tidal wind in the southern hemisphere.

Although no final conclusions regarding the phases of these harmonics can be

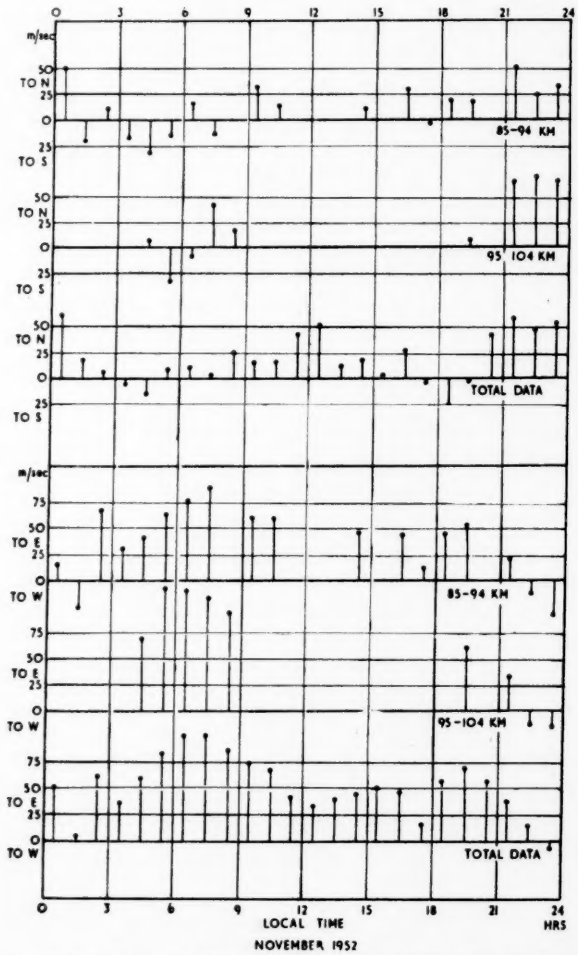


Fig. 8a. Comparison of the average diurnal variations of the N-S and E-W components of the winds for the three height groups during November

drawn from three months' observations only, the results given in Fig. 6 are consistent with the phase of the semi-diurnal tidal component derived from barometric oscillations, which in conjunction with tidal theory indicates that the wind at about 90 kms is directed towards the North at 0100 hours.

It is of interest to note that the phase of the 12 hour harmonic in October is about 2 hours later than either the November or December 12 hour harmonics.

VOL.
4
1953

It is also apparent from measurements made on individual days that the phase of the rotations can vary by as much as 3 hours from day to day.

5.3. *Wind gradient*

The results of § 4.5 are illustrated in Fig. 9 and they suggest a positive wind gradient of 3·6 m/sec/km. Gradients of this order and larger have been invoked to explain

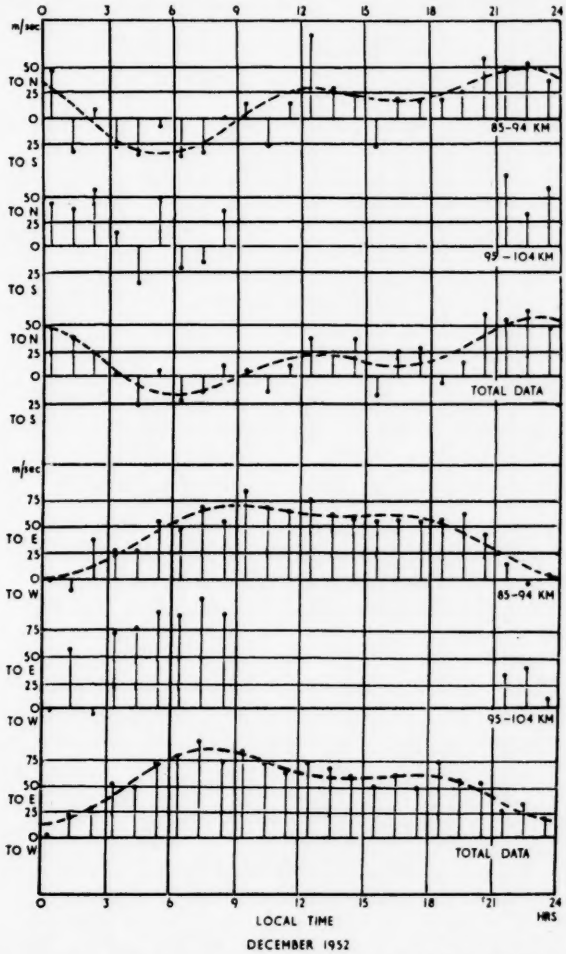


Fig. 8b. Comparison of the average diurnal variations of the N-S and E-W components of the winds for the three height groups during December.

the distortion of long enduring meteor trails (STÖRMER 1937), and radio measurements by GREENHOW (1952) on the relative velocity of two scattering centres of long enduring meteor trails give wind gradients varying between 1 and 7·5 m/sec/km. However. a wind gradient whose average value is that given above would indicate an average velocity of approximately 150 m/sec at 120 km. Such a value seems inconsistent with measurements made at 100–120 km using the fading technique, which

give values of the average wind velocity less than 100 m/sec. It appears therefore, that until further measurements are made, this value (3·6 m/sec/km) of the gradient should only be assigned to a small range of heights in the vicinity of 95 km.

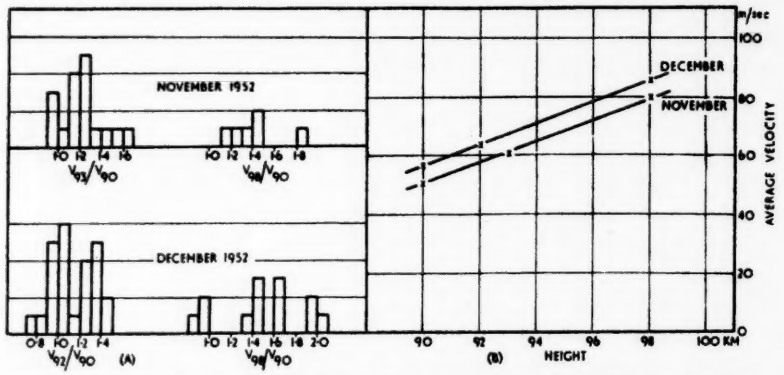


Fig. 9a. Distribution of the ratios of the average velocity at different heights.
Fig. 9b. Average velocity of the winds versus height.

6. Conclusions

It has been shown that radio observations of the drifts of meteor trails provide a reliable and continuous means of measuring upper atmospheric winds. The measurements refer directly to the real motion of the air at known heights between 80–105 km, and reveal the presence both of prevailing and of periodic horizontal winds.

It would appear that the results given in this paper are more definite and accurate than any previous measurements of winds in the upper atmosphere.

It is proposed to continue the work.

Acknowledgements—The authors wish to thank Professor L. G. H. Huxley for suggesting the field of work and for his constant encouragement and advice, and also Mr. D. T. Liddy, Mrs. D. A. Robertson and Messrs. R. N. Wheaton and C. R. McGee for their assistance in making the observations and reducing the recordings. They are also indebted to Mr. A. A. Weiss for helpful discussions.

This work was made possible through financial assistance from the University of Adelaide and the Radio Research Board of C.S.I.R.O. One of us (D.S.R.) is indebted to the *Imperial Chemical Industries (Australia and New Zealand) Ltd.* for the award of a Research Fellowship.

References

Beynon, W. J. G.	1948	Nature, **162**, 887
Ferrell, O. P.	1948	Proc. Inst. Radio Engrs. **36**, 879
Gerson, N. C.	1950	Nature, **166**, 316
Greenhow, J. S.	1952	J. Atmosph. Terr. Phys. **2**, 282
Manning, L. A., Villard, O. G., and Peterson, A. M.	1950	Proc. Inst. Radio Engrs. **38**, 877
Munro, G. H.	1950	Proc. Roy. Soc. A. **202**, 208
Phillips, G. J.	1952	J. Atmosph. Terr. Phys. **2**, 141
Salzberg, C. D., and Greenstone, R.	1951	J. Geophys. Res. **56**, 521
Störmer, C.	1937	Astrophysica Norvegica. **3**, No. 5

VOL. 4 1953

Journal of Atmospheric and Terrestrial Physics, 1954, Vol. 4, pp. 285 to 294. Pergamon Press Ltd., London

Excitation processes in the aurora and airglow

1. Absolute intensities, relative ultra-violet intensities and electron densities in high latitude aurorae.

M. J. Seaton

University College, London

(*Received* 2 *September* 1953)

ABSTRACT

The paper discusses certain auroral observations and seeks to obtain information in a form which is best suited to the study of physical processes.

An attempt is made to express largely descriptive information on the auroral brightness in terms of a quantity π (5577), the number of oxygen green line quanta emitted, in the line of sight, per cm^2 column per second. It is concluded that International Brightness Coefficients I, II, III and IV correspond to π (5577) being approximately equal to 2.10^9, 1.10^{10}, 1.10^{11} and 5.10^{12} cm^{-2} sec^{-1}. An extreme upper limit is estimated as 5.10^{15} cm^{-2} sec^{-1}.

Measured auroral intensities in ultra-violet vibration bands are compared with intensities obtained from theoretical and laboratory studies. By considering various v'' progressions (fixed upper state) it is shown that the auroral intensities are in error by an amount which increases systematically as the wavelength is decreased. An empirical correction curve is obtained.

Radio and optical evidence on the electron density is discussed and it is concluded that electron densities between 10^7 and 10^8 cm^{-3} occur in bright high latitude aurorae.

Introduction

Observations of Doppler-shifted H_α emissions in the auroral spectrum (Meinel, 1951; Vegard, 1952) give strong support for the view that the primary excitation mechanism of the aurora is provided by the entry of fast protons, possibly together with other positive ions, into the earth's atmosphere. Although the incident stream of charged particles must include sufficient electrons to ensure that it has no net electrical charge, the primary electrons will be stopped at great heights and will not give a large contribution to the auroral luminosity. Excitation by direct proton impact can, however, account for only a part of the auroral spectrum. In particular it cannot account for processes, such as the excitation of the forbidden atomic lines, which involve a change of electron spin. For such processes it is necessary to consider secondary excitation mechanisms.

Fast incident protons will loose energy in exciting, ionizing and dissociating the atmospheric constituents; further excitation can then arise from collisions with secondary electrons, from collisions involving excited particles and from recombination processes. When finally slowed down to velocities of the order of 1000 Km/sec, the protons capture electrons and become the fast hydrogen atoms which give the observed Balmer emissions.

Recent high dispersion studies (Petrie and Small, 1952; Vegard and Kvifte, 1952) have established with some certainty the identification of a great many spectral features. In order to assess the importance of various excitation processes it is also necessary to have information on the relative intensities, on the absolute emission rates and on the electron density. In the present paper the observational material is discussed and an attempt is made to obtain information in a form which is best suited to the study of physical processes. Since the aurora is un-

doubtedly a very complex phenomena, exhibiting many different forms and frequently undergoing rapid variations with time, we have in many cases to be content with obtaining sufficiently representative orders of magnitude for the quantities of interest in the construction of a physical theory.

The excitation of forbidden atomic lines in high latitude aurorae will be discussed in the second paper of the present series. Later papers will be concerned with excitation processes in low latitude aurorae and the airglow, and with the excitation of permitted atomic lines in high latitude aurorae.

2. ABSOLUTE BRIGHTNESS OF THE AURORA

The surface brightness of the aurora in a given wavelength λ is proportional to the number of photons of that wavelength emitted per second in a column in the line of sight of 1 cm^2 cross section. We denote this quantity by $\pi(\lambda)$. If $P(\lambda)$ is the average photon emission rate per cm^3 per sec and t the effective pathlength in Km, $\pi(\lambda)$ is given by $10^5\,tP(\lambda)$. The diameters of auroral rays are usually quoted to be of order 1 Km. For high latitude aurorae it appears reasonable to take t of order unity for all structures.

Although, so far as the author is aware, no exact quantitative information has been published on the absolute brightness of the aurora, it is not difficult to make order of magnitude estimates using descriptive information on the visual brightness. The apparent visual brightness may be expressed in terms of a quantity $\pi_v(5550)$, the equivalent number of photons in the wavelength of maximum visual sensitivity. It is readily shown that this is equal to $5{\cdot}6.10^{16}$ times the surface brightness in candles/cm^2. Since a high proportion of the visual brightness comes from the oxygen green line, λ 5577, we may take approximately $\pi\,(5577) = \frac{1}{2}\pi_v(5550)$. Using this relation we will express the absolute brightness in terms of $\pi\,(5577)$. If the relative intensities $I(\lambda)$ of various spectral features are known in arbitrary units the emission rates in the line of sight are readily obtained from

$$\pi(\lambda) = \frac{\lambda I(\lambda)}{5577\ I\ (5577)}\,\pi\,(5577) \qquad (1)$$

An extreme lower limit for the auroral brightness is provided by the brightness of the night sky, for which the results of recent determinations are summarized in Table 1; we adopt a figure of $\pi\,(5577) = 4.10^8$ cm^{-2} sec^{-1} for the night sky brightness.

Table 1. *Night sky brightness*

	$\pi\,(5577)$ *cm*$^{-2}$ *sec*$^{-1}$
ROACH and BARBIER (1950)	6.10^8
ROACH and PETTIT (1951)	$2\text{–}5.10^8$
HOUTGAST (1952)	$4{\cdot}4.10^8$

In Table 2 we give estimates of $\pi\,(5577)$ for aurorae of various International Brightness Coefficients (MITRA, 1947). For intensity I we use the data of ROACH

VOL. 4 1953

and PETTIT (1952) on the brightness of the Milky Way. The value for intensity II is interpolated. For intensity III we take the brightness of moonlit cumulus to be 2 micro-candles/cm² (DAUVILLIER, 1934). For intensity IV we take the surface brightness of the full moon to be 0·25 candles/cm² (HARDY and PERRIN, 1932) and assume that the solid angle subtended by the aurora is 10^3 times the solid angle of the full moon. The extreme upper limit, for which the surface brightness of the aurora equals that of the full moon, is only approached on very rare occasions (STØRMER, 1926; WURM, 1948).

Table 2. Absolute brightness of the auroral green line estimated from the apparent visual brightness

International Brightness Coefficient	*Description of brightness*	π (5577) $cm^{-2}\,sec^{-1}$
—	Extreme lower limit, night sky brightness	4.10^8
I	Brightness of Milky Way	2.10^9
II	Brightness of thin moonlit cirrus	1.10^{10}
III	Brightness of moonlit cumulus	1.10^{11}
IV	Illumination equals that of full moon	5.10^{12}
—	Extreme upper limit, brightness of full moon	5.10^{15}

A check on these order of magnitude estimates is provided by the following independent observations:

(1) It is well known that only the brightest aurorae give a visual colour sensation* (according to DAUVILLIER (1934) only those of intensity IV). The recent investigations of MIDDLETON and MAYO (1952) on the threshold brightness for colour vision show that π (5577) must be at least of order 10^{11}/cm²/sec for a green colour to be clearly seen.

(2) Using a photo-cell, DAUVILLIER (1934) found that intensity IV aurorae were 10^3 times as bright as intensity I aurorae.

(3) It will be shown in paper II of the present series that, for a pathlength of order 1 Km, the extreme upper limit of Table 2 is close to the maximum which is physically attainable, since it would require nearly all the atomic oxygen to be in excited states.

It is seen that the various estimates obtained give a self-consistent picture, and we can be reasonably certain that they provide sufficiently representative orders of magnitude for the discussion of excitation processes. It is, of course, very desirable that exact quantitative measurements should be made.

3. Relative Ultra-violet Intensities

The auroral spectrum is usually observed at angles of about 10° above the horizon. The resulting long pathlength through the earth's atmosphere leads to a great

* Since this was written Dr. N. HERLOFSON has commented that he would consider it more accurate to say that only the faintest aurorae do not give a visual colour sensation.

deal of absorption and scattering in the violet and ultra-violet spectral region, and in consequence to the possibility of large systematic errors in the determination of the relative intensities of ultra-violet features. Although most observers attempt to correct for atmospheric absorption, examination of the relative intensities of vibrational bands usually suggests that considerable errors remain.

The relative intensities $I(v', v'')$ in a v'' progression (fixed upper state v', variable v'') should be entirely independent of the type of source or excitation mechanism. We denote published auroral intensities by $I_O(v', v'')$ and estimates obtained from theoretical or laboratory studies by $I_T(v', v'')$. Systematic variations in the ratio $I_T(v', v'')/I_O(v', v'')$, which should be independent of v'', have previously been discussed by BATES (1949a) by BARBIER and WILLIAMS (1950) and by PETRIE and SMALL (1952b). These variations may be used to obtain an empirical correction factor to convert the published intensities to true relative emission intensities.

We consider the application of this method to the auroral intensities of VEGARD and KVIFTE (1945), which are the most complete set of auroral intensities at present available. For the I_T we take for the N_2^+ negative group the averaged theoretical and laboratory data quoted by BATES (1949a) and for the N_2 second positive group we use the theoretical results of NICHOLLS (1950).* In Fig. 1 we give plots of log (I_T/I_O) against wavelength, the v'' progressions for different values of v' being shifted by arbitrary amounts on the vertical scale. It is seen that each v'' progression gives a reasonably smooth and nearly linear function of λ, with similar slopes for the various values of v'. In Fig. 2 we combine these curves by shifting each progression on the vertical scale so as to obtain the smoothest combined curve.† Denoting the function obtained in Fig. 2 by f_λ, the corrected relative intensities are given by $I_c(\lambda) = f_\lambda I_O(\lambda)$. The scale has been adjusted so that extrapolation gives f_λ close to unity at λ 5577, which is usually taken as the reference wavelength for relative intensities. Essentially similar results may be obtained using the intensities of BARBIER and WILLIAMS (1950) and of PETRIE and SMALL (1952b).

The total atmospheric absorption may be calculated using the absorption coefficients given by BARBIER (1947). It is found that, for an elevation of 10° above the horizon, the total absorption correction obtained is not much greater than the correction f_λ of Fig. 2, despite the fact that VEGARD and KVIFTE have already made an absorption correction to their intensities. It should be borne in mind, however, that atmospheric absorption in the ultra-violet is critically dependent on variable atmospheric impurities and on the ozone content of the atmosphere. According to BATES (1954) the ozone content varies with latitude and season and is greater for high latitudes. It appears to be very desirable to

* For N_2^+ the theoretical and experimental I_T are in fair accord, but for the 2 P.G. of N_2 the laboratory I_T quoted by BATES are less reliable and the agreement between theory and experimental is less satisfactory. Use of the experimental I_T for the 2 P.G. gives results similar to those obtained with the intensities of NICHOLLS, but with a much greater scatter. The theoretical intensities are probably the more reliable, and for the present purpose have the additional advantage that they are unlikely to be in error by an amount which varies systematically with wavelength.

† The (2, 2) band of the N_2 2 P.G. at λ 3309, which gives an anomalous result in Fig. 1, has been omitted in Fig. 2. The anomaly is probably due to blending with other features (PETRIE and SMALL, 1952b).

VOL. 4 1953

make observations of atmospheric absorption at the same time and place as auroral observations. This could be done, for example, by means of observations on stars of known ultra-violet energy distribution.

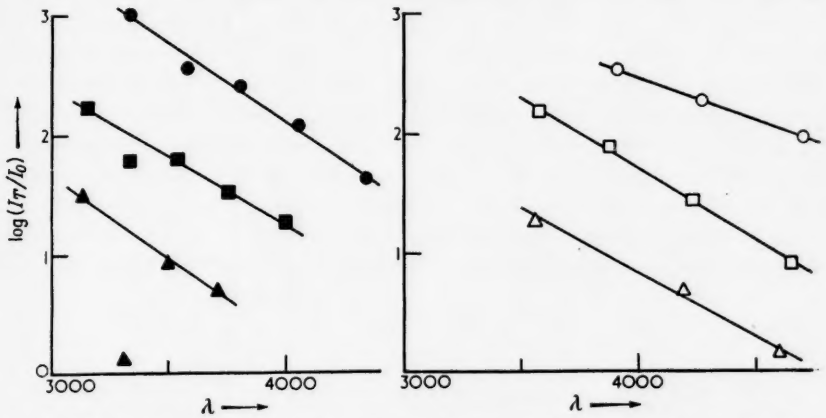


Fig. 1. Log (I_T/I_O) against wavelength, where the I_T are laboratory or theory intensities (see text) and the I_O are the measured auroral intensities of VEGARD and KVIFTE (1945)

● N_2 2 P.G. $v' = 0$ progression ○ N_2^+ N.G. $v' = 0$ progression
■ N_2 2 P.G. $v' = 1$ progression □ N_2^+ N.G. $v' = 1$ progression
▲ N_2 2 P.G. $v' = 2$ progression △ N_2^+ N.G. $v' = 2$ progression

The progressions for various upper states (v') are shifted by arbitrary amounts on the vertical scale.

Although the corrections obtained in Fig. 2 are very large, it is difficult to escape the conclusion that they arise from incomplete allowance for atmospheric absorption, possibly together with some instrumental error. The only alternative explanation likely to be suggested is that they arise from self-absorption. While

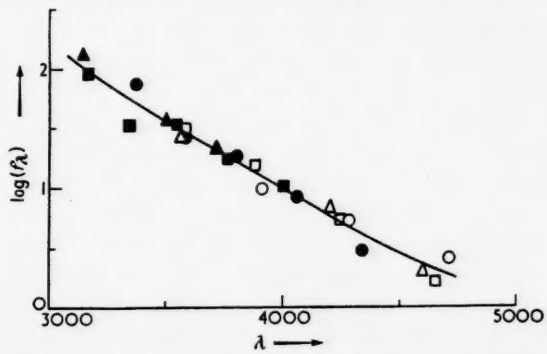


Fig. 2. Log (f_λ) against wavelength, obtained by shifting the various progressions of Fig. 1 on the vertical scale so as to obtain the smoothest combined curve. Corrected relative auroral intensities are given by

$$I_C(\lambda) = f_\lambda I_O(\lambda).$$

this would give a similar general trend, it is difficult to see how it could account for the detailed results, i.e. the similar variation of log (I_T/I_O) with λ for all progressions of both the N_2^+ N.G. and the N_2 2 P.G. It is nevertheless worth con-

sidering the condition for self-absorption to give observable intensity anomalies This can be done without difficulty for the N_2^+ negative group.

The condition for self-absorption to begin to be important may be taken to be that the total optical depth in the line of sight should be unity for the centre of the strongest line. It is necessary to take the rotational structure of the bands into account, but since we are only interested in orders of magnitude we neglect such effects as spin and hyperfine structure and the two to one alternation of the statistical weights of the rotational levels. We use the usual expressions for the relative population of the rotation level giving the strongest absorption, for the rotational line strength, and for the absorption coefficient at the centre of a Doppler-broadened line (HERZBERG, 1950; MITCHELL and ZEMANSKY, 1934). Using the f-values of SHULL (1951) and the fractional transition probabilities of BATES (1949a) we obtain finally for the central optical depth of the strongest line

$$\tau = \frac{2.10^{-11}}{T} 10^5 t\bar{n}(N_2^+, X\,^2\Sigma, v'' = 0) \tag{2}$$

where $\bar{n}(N_2^+, X^2\Sigma, v'' = 0)$ is the average number of N_2^+ ions per cm^3 in the $X\,^2\Sigma$, $v'' = 0$ state. Putting $\tau = 1$, $t = 1$ Km, and $T = 500°$K, this gives

$$\bar{n}(N_2^+) > \bar{n}(N_2^+, X^2\Sigma, v'' = 0) = 4.10^8 \text{ cm}^{-3} \tag{3}$$

as the condition for self-absorption to begin to be important. This will be further discussed in section 4.3.

VOL.
4
1953

4. ELECTRON DENSITIES

4.1. *The radio evidence*

During recent years radar echoes from the aurora have been studied by a number of workers; frequencies on which echoes have been obtained are summarized in Table 3. Both discrete and diffuse echoes are observed, the former being ascribed to auroral rays and the latter to extended irregular surfaces. Unfortunately there has been considerable controversy about the correct method of estimating electron densities from these results.

If the echoes arise from ionosphere-type total reflection, a lower limit for $n(e)$ in cm^{-3} is given by $1.24.10^4 f^2$, where f is a frequency in Mc/s on which well defined echoes are obtained. Table 3 shows that this mechanism would require $n(e)$ to be at least 10^7–10^8 cm^{-3} in high latitude aurorae.

An alternative view (HERLOFSON, 1947) is that the electron density is below the critical value and that the echoes result from partial reflection due to variations in refractive index. For this mechanism further information on the magnitude of $n(e)$ can only be obtained by measuring the returned power and making some assumptions about the space distribution of the electrons.

ASPINALL and HAWKINS (1950) assumed that their discrete echoes arose from partial reflection due to auroral rays, which were assumed to be cylinders of 1 Km diameter containing a homogeneous electron density. From a measurement of returned power they obtained $n(e) \simeq 6.10^5$/cm^3 from their observations and $n(e) \simeq 9.10^5$/cm^{-3} from the observations of LOVELL, CLEGG and ELLYETT (1947). The proposed partial reflection mechanism requires a sharp boundary for the

auroral ionization. According to CHAPMAN (1952), HAWKINS now considers this to be improbable, due to the rapid diffusion to be expected at auroral levels. HAWKINS and his colleagues now regard the total reflection mechanism as more probable, giving $n(e) \geqslant 6{\cdot}5.10^7$ cm^{-3}.

Table 3. Summary of auroral radar soundings

Authors	*Station*	*Frequency (Mc/s)*	
		Echoes	*No echoes*
LOVELL, CLEGG and ELLYETT (1947)	Manchester, England	46	72
ASPINALL and HAWKINS (1950)	Manchester, England	72	
FORSYTH, PETRIE, VAWTER and CURRIE (1950)	Saskatoon, Canada	106·5	3000
GERSON (1951)	Radio amateurs in U.S.A., reflections from aurorae in Canada	50	
DAVIDSON (1951)	Concord, Mass., U.S.A.	17·3	
PIERCE (1947)	Cambridge, Mass., U.S.A.	3·5	
HELLGREN and MEOS (1952)	Kiruna, Sweden	30·3	
MCKINLEY and MILLMAN (1953)	Ottawa, Canada	33	

HELLGREN and MEOS (1952) have carried out continuous radar observations of high latitude aurorae on a frequency of 30·3 Mc/s during the period May 1951–March 1952. Using a method similar to the original method of ASPINALL and HAWKINS they estimate $n(e)$ to be 2.10^6 cm^{-3} for the rays. No mention is made of a correction for absorption in the measurement of the returned power, although the close correlation between aurorae and polar blackouts suggests that such a correction may be important. The diffusion difficulty mentioned in the previous paragraph is also applicable to this case.

It has been seen (section 2) that the visual brightness of the aurora varies between wide limits. Even if a value of 2.10^6/cm^3 were accepted as an average value for $n(e)$, it would appear to be very improbable that the critical value ($1{\cdot}1.10^7$/cm^3) was never reached during the period of these observations. However, if the majority of echoes were produced by partial reflection, greatly increased values of the returned power would be expected whenever the critical densities were reached. The fact that no echoes of greatly increased intensity were reported favours the view that all the echoes were produced by total reflection.

It should be possible to distinguish between partial and total reflection by making dual frequency measurements of the ratio between the returned and transmitted power. Dr. P. A. FORSYTH, of the University of Saskatchewan, has informed the writer of some measurements of this type that he has carried out

using frequencies of 56 and 106·5 Mc/s. He states that, although the measurements cannot be interpreted in terms of total reflection from a simple surface, they certainly do not indicate a partial reflection mechanism. The correct interpretation appears to be that electron densities greater than the critical values are present within limited volumes. The 56 Mc/s echoes are observed from most auroral displays but the 106·5 Mc/s echoes are much less frequent and of shorter duration. Electron densities of the order of the critical densities, $4.10^7/\text{cm}^3$ and $1{\cdot}4.10^8/\text{cm}^3$, can apparently occur only near the 100 Km level.*

4.2. *The optical evidence*

There is little doubt that the N_2^+ negative group is emitted after simultaneous ionization and excitation of N_2 (BATES, 1949a).† A lower limit for the total electron production rate is then given by the rate of emission of negative group quanta, P(N.G.). About one half of the N.G. quanta are emitted in the λ 3914 band, so that we may put $P(\text{N.G.}) \simeq 2\pi\,(3914)/10^5 t$. VEGARD and KVIFTE (1945) give the intensity of λ 3914 to be about half that of λ 5577. Introducing the ultraviolet correction factor f_λ of section 2, this gives $P(\text{N.G.}) \simeq f_{3914}\pi(5577)/10^5 t$. Equating the electron production rate to the recombination rate we obtain finally

$$n(e) > \left[\frac{f_{3914}\pi\,(5577)}{10^5 t\alpha^{\text{rec}}}\right]^{\frac{1}{2}}, \tag{4}$$

where α^{rec} is the recombination coefficient.

In this expression we substitute $\pi\,(5577) = 5.10^{10}\ \text{cm}^{-2}\,\text{sec}^{-1}$, which is unlikely to be an over-estimate for a bright display, $f_{3914} = 10$, $t = 1$ and $\alpha^{\text{rec}} = 10^{-8}$ $\text{cm}^3\,\text{sec}^{-1}$ (a typical E-layer value (RYDBECK, 1946)). This gives $n(e) > 2.10^7\ \text{cm}^{-3}$. Despite the various uncertainties‡ it would appear to be improbable that $n(e)$ could be much less than $10^7\ \text{cm}^{-3}$ for a bright display, since we have neglected all ionization processes other than production of $N_2^+(B\,^2\Sigma)$, and have assumed a comparatively low value for $\pi\,(5577)$.

A second estimate may be made by considering the enhanced N_2^+ emission due to resonance in sunlit aurorae. Since the intensity of solar radiation and the transition probabilities are known, it is possible to estimate $n(N_2^+)$ from the emission rate, and hence to obtain a lower limit for $n(e)$. High sunlit rays are usually described as having a faint greyish colour, although the spectroscope shows much of the light to be in the blue-violet N_2^+ bands. Occasionally brighter sunlit rays are described as having a distinct blue colour (STØRMER, 1926, 1929, 1938; RAYLEIGH, 1922). Assuming this to be a true colour sensation, the effective number of quanta in the negative group can scarcely be much less than $10^{11}/\text{cm}^2/\text{sec}$, which, allowing for atmospheric absorption, corresponds to $\pi\,(\text{N.G.}) \simeq 10^{12}$, $\pi\,(3914) \simeq 5.10^{11}$ and $P(3914) \simeq 5.10^6\ \text{cm}^{-3}\,\text{sec}^{-1}$ (for $t = 1$ Km). The results of BATES (1949b), corrected for the improved transition probabilities of SHULL

* FORSYTH's work has been published (J. Geophys. Res. 1953 **58**, 53) since the present paper was written.

† The evidence for this will be discussed further in paper II of the present series.

‡ The most serious uncertainties are probably the value of α^{rec}, since auroral conditions may differ from those in the normal E-layer, and the value of the effective pathlength t.

VOL. 4 1953

(1951), give for resonant emission $P(3914) \simeq 0{\cdot}3n(N_2^+)$, so that we obtain $n(N_2^+) \simeq 2.10^7$ cm^{-3}. Since the negative ion to electron ratio will certainly be small in sunlight (BATES and MASSEY, 1951), we may take $n(N_2^+)$ as a lower limit for $n(e)$, giving $n(e) \geqslant 2.10^7$ cm^{-3}.

This estimate does not provide any direct information on the value of $n(e)$ appropriate to auroral forms not in the sunlit atmosphere. It is of interest, however, that MCKINLEY and MILLMAN (1953) obtained echoes, on a frequency of 33 Mc/s, from aurorae described as "violet sunlit rays." If the brightness of these was above the colour vision threshold,* it would appear to be very probable that the electron density exceeded the critical value ($1{\cdot}4.10^7$ cm^{-3}). In this case it would also appear to be very probable that all other echoes obtained by MCKINLEY and MILLMAN were due to total reflection.

4.3. *Conclusion*

It may be concluded that the available evidence favours the view that electron densities between 10^7 and 10^8 cm^{-3} occur in bright high latitude aurorae.

In discussing the optical evidence we have used the ultra-violet correction factor f_λ of section 3. It should be noted that if the anomalies in the band intensities were attributed to self-absorption effects, as in section 3, it would be necessary to assume even greater values for $n(e)$.

Acknowledgments—The author wishes to thank Professor H. S. W. MASSEY, F.R.S., Professor D. R. BATES, Dr. W. PETRIE, Mr. W. R. PIGGOTT and Dr. T. R. KAISER for many helpful discussions and comments on this work. He is also particularly indebted to Dr. P. A. FORSYTH for information on the results of his radar observations in advance of publication.

REFERENCES

ASPINALL, A., and HAWKINS, G. S.	1950	J. Brit. Astr. Ass. **60,** 130
BARBIER, D.	1947	Ann. d'Astrophys. **10,** 141
BARBIER, D., and WILLIAMS, D. R.	1950	J. Geophys. Res. **55,** 401
BATES, D. R.	1949*a*	Proc. Roy. Soc. A. **196,** 217
	1949*b*	Proc. Roy. Soc. A **196,** 562
	1954	The Physics of the Upper Atmosphere, Chapter 12 of "The Solar System" (Ed. G. P. KUIPER), Vol. 2, University of Chicago Press
BATES, D. R., and MASSEY, H. S. W.	1951	J. Atmosph. Terr. Phys. **2,** 1
CHAPMAN, S.	1952	J. Atmosph. Terr. Phys. **3,** 1
DAUVILLIER, M. A.	1934	Jour. de Phys. et Radium **5,** 398
DAVIDSON, D.	1951	Nature **167,** 277
FORSYTH, P. A., PETRIE, W., VAWTER, F., and CURRIE, B. W.	1950	Nature **165,** 561
GERSON, N. C.	1951	Nature **167,** 804
HARDY, A. C., and PERRIN, F. H.	1932	The Principles of Optics (1st edition), McGraw-Hill, New York, 167

* A surface with brightness below the colour threshold is frequently described as "blue-grey." It cannot therefore be concluded with certainty that a faint surface described as "violet" has a brightness in excess of the colour threshold.

Hellgren, G., and Meos, J.	1952	Tellus **3,** 249
Herlofson, N.	1947	Nature **160,** 867
Herzberg, G.	1950	Spectra of diatomic molecules (2nd edition), Van Nostrand, New York, Chs. III and IV.
Houtgast, J.	1952	Mém. Soc. Roy. Liège **12,** Fsc. I–II, 117
Lovell, A. C. B., Clegg, J. A., and Ellyett, C. D.	1947	Nature **160,** 372
McKinley, D. W. R., and Millman, P. M.	1953	Canadian J. Phys. **31,** 171
Meinel, A. B.	1951	Rep. Prog. Phys. **14,** 121
Middleton, W. E. K., and Mayo, E. G.	1952	J. Opt. Soc. Amer. **42,** 116
Mitchell, A. C. G., and Zemansky, M. W.	1934	"Resonance Radiation and Excited Atoms" (Cambridge University Press), Ch. III
Mitra, S. K.	1947	The Upper Atmosphere (1st edition). The Asiatic Society, Calcutta, 496
Nicholls, R. W.	1950	Phys. Rev. **77,** 421
Petrie, W., and Small, R.	1952*a*	Astrophys. J. **116,** 433
	1952*b*	J. Geophys. Res. **57,** 51
Pierce, J. A.	1947	Phys. Rev. **71,** 88
Rayleigh, Lord	1922	Proc. Roy. Soc. A **101,** 114
Roach, F. E., and Barbier, D.	1950	Trans. Amer. Geophys. Union **31,** 7
Roach, F. E., and Pettit, H. B.	1951	J. Geophys. Res. **56,** 325
	1952	Mém. Soc. Roy. Liège **12,** Fsc. I–II, 13
Rydbeck, O. E. H.	1946	Trans. Chalmers Univ. of Technology **53**
Shull, H.	1951	Astrophys. J. **114,** 546
Størmer, C.	1926	Geofys. Publ. **4,** No. 7
	1929	Nature **123,** 868
	1938	Nature **142,** 1034
Vegard, L.	1952	Geofys. Publ. **18,** No. 5
Vegard, L., and Kvifte, G.	1945	Geofys. Publ. **16,** No. 7
	1952	Geofys. Publ. **18,** No. 3
Wurm, K.	1948	Z. Astrophys. **25,** 28

VOL. 4 1953

Journal of Atmospheric and Terrestrial Physics, 1954, Vol. 4, pp. 295 to 313. Pergamon Press Ltd., London

Excitation processes in the aurora and airglow

2. Excitation of forbidden atomic lines in high latitude aurorae

M. J. SEATON

University College, London

(*Received* 8 *September* 1953)

ABSTRACT

A summary is given of the various excitation and deactivation processes which may occur, and an attempt is made to decide which of these will be of major importance.

Relative intensities of the [O I] lines (λ 5577, and 6300 + 6364) show that deactivation must occur. If electron deactivation alone occurs, $n(e)$ must be $\geqslant 7.10^7/cm^3$ at the lower border and $\geqslant 2.10^7/cm^3$ at the upper border. It was concluded in paper 1 that such values are possible for a bright display. Various deactivation processes involving collisions with gas molecules are discussed. Although none of these appear to be particularly promising, there are several processes which may contribute.

At and above the level of maximum luminosity it is considered that the [O I] lines are excited by electron impact. Below this level there is some evidence that an additional excitation mechanism becomes important. This is probably $O_2^+ + e \rightarrow O' + O''$.

Electron deactivation can account for the relative intensities of the [N I] lines (λ 3467 and 5199) with electron densities similar to those required for the [O I] lines. An improbably high degree of nitrogen dissociation would be required if electron impact on atomic nitrogen was assumed to be responsible for excitation of the [N I] lines. The forbidden line emission rate resulting from $N_2^+ + e \rightarrow N' + N''$ is estimated and found to be in good agreement with observation.

Proposed identifications with [N II], [O II] and [O III] lines are discussed. The only one for which the identification does not appear to be in doubt, [N II] λ 5755, is probably excited by dissociative ionization of N_2, $N_2 + X \rightarrow N' + N^{+''} + e + X$.

1. INTRODUCTION

The forbidden lines of O I, λ 5577 and 6300 + 6364, have long been known as prominent features in the spectrum of high latitude aurorae, and more recent work (BERNARD, 1941; PETRIE and SMALL, 1952a; GARTLEIN and SHERMAN, 1952) has also shown that the forbidden lines of N I, λ 3467 and 5199, are regularly excited. Numerous identifications with forbidden lines of N II, O II and O III have also been suggested, but most of these appear to be very doubtful.

Using the results obtained in paper 1 (SEATON, 1954a) of the present series and of recent work on collision processes, we discuss in the present paper the various excitation and deactivation mechanism which are likely to determine the forbidden line intensities. In paper 1 we briefly discussed the general view that the primary excitation processes of the aurorae were due to the entry of fast protons, possibly together with other positive ions, into the Earth's atmosphere. At the heights of auroral phenomena the atmosphere consists largely of molecular nitrogen and atomic oxygen, possibly together with some atomic nitrogen and with molecular oxygen becoming important at the lower levels (~100 Km) (BATES, 1954; BATES and WITHERSPOON, 1952). There can be little doubt that most of the forbidden atomic emissions arise from various secondary excitation processes and not from excitation of the main atmospheric constitutions by direct proton impact. Before discussing these secondary processes it is convenient to summaries the main processes expected to be responsible for stopping the fast protons.

Dissociation of N_2 by electron impact appears to have a very low efficiency (MASSEY and BURHOP, 1952), and the same is probably true for dissociation by proton impact. Collisional ionization,

$$N_2 + H^+ \rightarrow N_2^{+\prime} + H^+ + e \qquad (1)$$

is likely to be much more important. A considerable proportion of the N_2^+ formed by (1) will probably be in excited states with considerable vibrational development. Thus MEINEL and FAN (1952) have obtained strong N_2^+ emissions on bombarding molecular nitrogen with high speed protons, and found the vibrational development to be similar to that observed in the aurora. It is also of interest that MEINEL (1952) reports a striking correlation between the intensity of the N_2^+ auroral emissions and the auroral H_α emission.

Since (1) apparently gives considerable excitation to N_2^+, we may also expect the dissociative ionization reaction,

$$N_2 + H^+ \rightarrow N' + N^{+\prime\prime} + H^+ + e \qquad (2)$$

to be of importance (it being a limiting form of (1)). Charge exchange reactions, responsible for producing the excited hydrogen atoms giving the observed Balmer emissions, are unlikely to be important as secondary excitation processes for the forbidden lines, since these reactions only occur when the protons have lost some 99% of their initial energy (MEINEL, 1952).

The mean initial kinetic energies of the electrons formed by reactions such as (1) may be expected to be of order 10—20 eV.

In both the present and the previous paper it is assumed that the N_2^+ auroral emissions result from processes such as (1) giving simultaneous ionization and excitation of N_2. This view has been criticized by SWINGS (1948), who considers collisional excitation of N_2^+ to be a more probable mechanism. Our main objection to this suggestion is that N_2^+ is a rare constituent, its density being unlikely to exceed 10^{-4} times that of the main atmospheric gases, and it is therefore difficult to see how collisions with N_2^+ could give rise to some of the most intense emissions in the whole spectrum. Thus, for example, it is readily shown that the intensity resulting from electron excitation of N_2^+ must be many orders of magnitude less than the intensity of the forbidden oxygen lines, since the N_2^+ electron excitation cross sections cannot greatly exceed the oxygen cross sections (which are of order gas kinetic). A further objection is that collisional excitation of N_2^+ is unlikely to proceed more rapidly than recombination (see section 2.5) and therefore a very high rate of production of N_2^+ would be required for the mechanism proposed by SWINGS. There are, however, no grounds for assuming that ionization of N_2 would give N_2^+ exclusively in the ground state.

In discussing excitation processes we will assume that the emission rate of N_2^+ negative group quanta gives a lower limit for the ionization rate (the total ionization rate is probably greater than the N.G. emission rate by a factor of order 2 or 3). With this assumption we may then estimate the relative intensities resulting from various processes and compare our results with observations. An

VOL. 4 1953

alternative approach would be to attempt to calculate the absolute emission rates.* A disadvantage of this approach would be that such calculations would be more uncertain and also that the observed absolute emission rate are known less reliably than the relative intensities. Occasionally, however, attempts to estimate absolute emission rates are unavoidable.

Throughout the present paper we use the notation introduced in paper 1. $P(\lambda)$ denotes the average photon emission rate per cm^3 per sec, and $\pi(\lambda)$ the photon emission rate per cm^2 column per sec in the line of sight. These are related by $\pi(\lambda) = 10^5 t P(\lambda)$, where t is the effective pathlength in Km (assumed to be of order unity). Intensities are denoted by $I(\lambda)$ and are given relative to I (5577) = 100 unless the contrary is stated. The ultra-violet correction factor of paper 1 will be used in the present paper, with the exception of section 4 where we consider intensities of λ 3467 both with and without this factor.

2. Collision Processes

2.1. The forbidden atomic lines arise from transitions between the three ground configuration terms of the atoms concerned. It will occasionally be found convenient to denote these terms by subscripts $n = 1, 2, 3$ in order of increasing excitation energy. Table 1 gives, for O I and N I, the excitation energies E_n, the statistical weights ω_n, the wavelengths λ_{nm} and the transition probabilities A_{nm}. The values of Garstang (1951, 1952) have been used for the A_{nm} of O I and for the ^{2}D $\rightarrow$ ^{4}S transition of N I, and the A_{nm} of Pasternack (1940), corrected for the improved S_q integrals of Garstang (1952), have been used for the N I $^2P \rightarrow {}^4S$ and $^2P \rightarrow {}^2D$ transitions. In all cases the A_{nm} have been averaged over initial J-levels and summed over final J-levels.

Table 1

Atom	*Term*	n	ω_n	E_n (*eV*)	*Atom*	*Transition* $n \rightarrow m$	λ_{nm}(Å)	A_{nm} (sec^{-1})
O I	^1S	3	1	4·17	O I	$^1S \rightarrow {}^3P$	2972, 2958	0·078
	^1D	2	5	1·96		$^1S \rightarrow {}^1D$	5577	1·28
	^3P	1	9	0·00		$^1D \rightarrow {}^3P$	6300, 6364	0·0091
N I	^2P	3	6	3·56	N I	$^2P \rightarrow {}^4S$	3466·5	$5{\cdot}4.10^{-3}$
	^2D	2	10	2·37		$^2P \rightarrow {}^2D$	10407, 10398	0·079
	^4S	1	4	0·00		$^2D \rightarrow {}^4S$	5198, 5200	$1{\cdot}04.10^{-5}$

2.2. *Notation.* Denoting the number of atoms A per cm^3 by $n(A)$, the number of collisions of the type

$$A + B \rightarrow A' + B' \qquad (3)$$

* It may be recalled that in paper 1 electron densities were calculated from estimates of the N.G. emission rate, which was assumed to give a lower limit for the ionization rate. The values obtained were in reasonable agreement with estimates obtained from other evidence. This provides some confidence in the consistency of our assumptions.

occurring per sec per cm^3 is denoted by $\alpha_{(3)}n(A)n(B)$, the subscript giving the equation number. An exception is made for inelastic electron collisions,

$$X_n + e \rightarrow X_m + e \tag{4}$$

for which the rate coefficient is written as α^e_{nm}.

In discussing the possible importance of certain deactivation processes it is convenient to introduce the quantity γ, the transition probability per gas kinetic collision. Within the range of gas kinetic cross sections and temperatures occurring in the upper atmosphere we may take $\alpha \simeq 3.10^{-10}\gamma$.

2.3. *The equations of equilibrium.* The equations of equilibrium, assuming the only important processes to be electron collisions with atoms and radiative transitions, have been discussed by BARBIER (1948) and by SEATON (1954b). Due to the great complexity to be expected in auroral processes a somewhat different approach is adopted, involving no explicit assumptions about the types of collisions which may occur. We introduce the following notation:

S_3 = the total number of atoms entering state 3 per sec per cm^3

S_2 = the total number of atoms entering state 2 per sec per cm^3

by all processes other than cascading or deactivation from state 3.

d_n = the total probability per sec of an atom in state n suffering collisional deactivation

d_{32} = the probability per sec of an atom in state 3 being deactivated by collisions to state 2 (d_3 includes d_{32}).

In this notation the equations of equilibrium take the form

$$(A_{32} + A_{31} + d_3)n(X_3) = S_3 \tag{5}$$

$$(A_{21} + d_2)n(X_2) = S_2 + (A_{32} + d_{32})n(X_3) \tag{6}$$

To derive the ratio $n(X_2)/n(X_3)$ we make the assumption that S_3 does not involve $n(X_2)$, i.e. we neglect $2 \rightarrow 3$ transitions. This is always justified if $n(X_2)$ is sufficiently small. With this assumption we obtain

$$\frac{n(X_2)}{n(X_3)} = \left(\frac{A_{31} + A_{32} + d_3}{A_{21} + d_2}\right)\left[\frac{S_2}{S_3} + \frac{A_{32} + d_{32}}{A_{31} + A_{32} + d_3}\right] \tag{7}$$

The quantum emission rates are given by $P(\lambda_{nm}) = A_{nm}n(Xm)$ and the intensities are given by $I(\lambda_{nm}) = (E_n - E_m)P(\lambda_{nm}) \times$ constant.

The transition probabilities A_{31} and A_{32} are in all cases much greater than A_{21}. We therefore consider the further approximation of neglecting d_3 and d_{32} but retaining d_2. Substituting numerical values for the energies and transition probabilities we obtain for the two main cases of interest

O I: $$\frac{I(6300 + 64)}{I(5577)} = \frac{0{\cdot}94}{1 + 110d_2}\left[\frac{S_2}{S_3} + 0{\cdot}94\right] \tag{8}$$

and

N I: $$\frac{I(5199)}{I(3467)} = \frac{10}{1 + 9{\cdot}6.10^4 d_2}\left[\frac{S_2}{S_3} + 0{\cdot}94\right] \tag{9}$$

VOL. 4 1953

In this approximation the relative intensities are seen to be determined by the two parameters d_2 and S_2/S_3. In the absence of deactivation ($d_2 = 0$), $I(6300 + 64)/I(5577)$ is greater than 0·88 and $I(5199)/I(3467)$ is greater than 9·4.

The neglect of d_3 and d_{32} is probably justified in all cases for O I, but may not be justified for N I (see sections 3.3 and 4.2).

2.4. *Electron excitation.* Cross sections for electron excitation of O I have been calculated by SEATON (1953a). The calculated deactivation coefficients $\alpha^e_{nm}(E_n > E_m)$ are given in Fig. 1 as functions of electron temperature T_e(°K). The corresponding excitation coefficients are given by

$$\alpha^e_{mn} = \frac{\omega_n}{\omega_m}\,\alpha^e_{nm} e^{-(E_n - E_m)/kT_e} \tag{10}$$

Detailed calculations of the cross sections for N I have not been made. However, it is known (SEATON, 1953b) that, apart from external symmetry factors, the

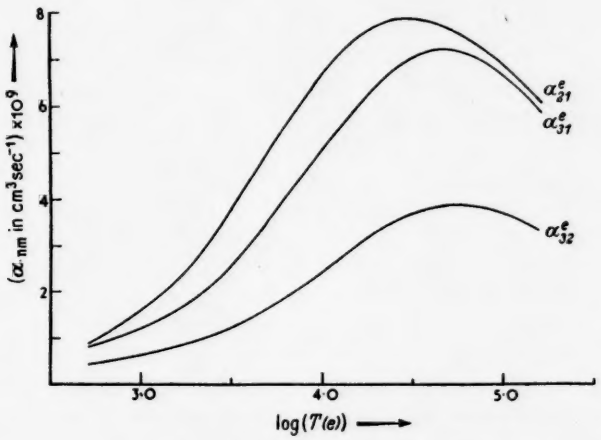


Fig. 1. Electron deactivation coefficients for O I as functions of electron temperature. Subscripts 1, 2 and 3 for $2p^4\ ^3P$, ^1D and ^1S.

magnitudes of the cross sections are similar for similar ions. Allowing for the external factors we may take

$$\alpha^e_{21}(\text{N I}) = 0{\cdot}52\alpha^e_{21}(\text{O I}),\quad \alpha^e_{31}(\text{N I}) = 0{\cdot}52\alpha^e_{31}(\text{O I}),\quad \alpha^e_{32}(\text{N I}) = 2{\cdot}3\alpha^e_{32}(\text{O I}) \tag{11}$$

It is probable that the values of α^e_{21}(O I) are correct to within $\pm 30\%$ and that the other O I coefficients are not in error by a factor greater than 2. The N I coefficients obtained from (11) are probably correct to within a factor of order two.

2.5. *Dissociative recombination.* The recombination reactions

$$O_2^+ + e \rightarrow O' + O'' \tag{12}$$

and

$$N_2^+ + e \rightarrow N' + N'' \tag{13}$$

may be of considerable importance as secondary excitation processes (BATES, MASSEY and PEARSE, 1948). The energy available for excitation of the atomic fragments is the difference between the ionization potential and the dissociation energy of the neutral molecule. For O_2 this energy is 7·1 eV, which is sufficient to give on the right hand side of (12) any combination of the ground configuration terms except $^1S + {}^1S$.* For N_2 the energy available for excitation is 5·82 eV using the dissociation energy advocated by GAYDON (1952) (9·756 eV) and is 8·20 eV using the dissociation energy advocated by HERZBERG (1950) (7·373 eV). The most recent evidence (GAYDON, 1952; DOUGLAS, 1952) favours the former value but the possibility of the second value being correct cannot be completely discounted. With GAYDON'S value (13) can give any combination of the ground terms except $^2D + {}^2D$ and $^2D + {}^2P$. For the latter case the energy defect is only 0·11 eV, and could possibly be made up by kinetic or vibrational energy (MITRA, 1951).

The available experimental evidence on the efficiency of dissociative recombination, and its theoretical interpretation, has been discussed in a recent review by MASSEY (1952). The rate coefficients $\alpha_{(12)}$ and $\alpha_{(13)}$ are probably large, between 10^{-8} and 10^{-6} cm^3/sec. In discussing these reactions as excitation mechanisms it is also important to estimate the S_2/S_3 ratio. Previous estimates (MITRA, 1951) have been largely based on considerations of energy resonance. However in this, as in many other upper atmosphere problems, the idea that close energy resonance gives large coefficients has probably been used rather indiscriminately. Although such considerations are of undoubted relevance in discussing certain processes involving only a transfer of electronic energy, they are not applicable for many molecular processes (MASSEY and BURHOP, 1952; OLDENBERG, 1952). For dissociative recombination it is necessary to consider the course of the reaction in more detail. We may consider that the electron first attaches itself to the molecular ion to give a repulsive state of the neutral molecule, which then breaks up into atomic fragments. To make a reliable estimate of S_2/S_3 it would be necessary to have more information than is at present available on the repulsive potential energy curves. However, it is probable that a number of such curves contribute, in which case it is unlikely that there would be great selectivity in producing atoms in one excited state in preference to the other. In the absence of further information the most reasonable value to adopt for the S_2/S_3 ratio would appear to be the statistical weight ratio, ω_2/ω_3.

2.7. *Selective excitation of state* 3. In discussing the interpretation of the auroral spectrum in sections 3.3 and 4.2 we will consider, for the sake of completeness, the lower limit of zero for the S_2/S_3 ratio, although there is no obvious example of a mechanism likely to give this value. The reaction

$$N_2^+ + O^- \rightarrow N_2(B\,^3\Pi,\, v' = 9) + O(^1S) \qquad (14)$$

has frequently been discussed in connection with auroral excitation (MITRA, 1946; GHOSH, 1946; NICOLET, 1945), but BATES, MASSEY and PEARSE (1948) have shown

* If the O_2^+ in (12) is in the metastable $a\,^4\Pi_u$ state, the energy available for excitation is 11·0 eV. This is sufficient for excitation of O I $3p\,^5P$ (requiring 10·69 eV) and O I $3p\,^3P$ (requiring 10·94 eV).

VOL. 4 1953

that it is unlikely to be of importance. It is in any case very doubtful whether a reaction of this type would give great selectivity in exciting $O\,^1S$ (OLDENBERG, 1952).

2.8. *Deactivation processes.* We consider the various mechanisms by which $O(^1D)$ could be deactivated in collisions with gas molecules. For

$$O(^1S) + O_2 \rightarrow O(^3P \text{ or } ^1D) + O_2(Y) \tag{15}$$

VEGARD and KVIFTE (1947) give $\gamma_{(15)} = 3{\cdot}6.10^{-6}$ and for

$$O(^1D) + O_2 \rightarrow O(^3P) + O_2(Y) \tag{16}$$

they give $\gamma_{(16)}$ of the same order as $\gamma_{(15)}$. Some of the difficulties in the experimental method used have been discussed by SAYERS and EMELÉUS (1952).

The value obtained by VEGARD and KVIFTE are surprisingly small, since in both cases part of the energy can be taken up as electronic energy of O_2 (the state Y being $a\,^1\Delta_g$ or $b\,^1\Sigma_g^+$), and although electron exchange is required, there is no violation of the WIGNER spin conservation rule (MASSEY and BURHOP, 1952). If resonance considerations have any relevance in discussing these reactions, as appears possible (MASSEY and BURHOP, 1952; OLDENBERG, 1952), one would expect $\gamma_{(16)}$ to be considerably greater* than $\gamma_{(15)}$. It thus appears possible that $\gamma_{(16)}$ *might* be considerably greater than $3{\cdot}6.10^{-6}$.

For

$$O(^1D) + N_2 \rightarrow O(^3P) + N_2' \tag{17}$$

electronic excitation of N_2 is energetically impossible and vibrational excitation is unlikely to be an efficient process, particularly in view of the fact that a violation of the spin conservation rule is involved. The value of $\gamma_{(17)}$ would therefore be expected to be very small.†

The radiative reactions

$$O(^1D) + O_2 \rightarrow O_3 + h\nu \tag{18}$$

and

$$O(^1D) + N_2 \rightarrow N_2O + h\nu \tag{19}$$

would be expected to have cross sections much less than gas kinetic. Some idea of the order of magnitude of their probabilities may be estimated from the cross sections for the inverse processes, but these estimates are necessarily crude since different initial and final states may be involved. However a rough calculation indicates that the value of $\gamma_{(18)}$ will be considerably less than 10^{-6} and that $\gamma_{(19)}$ will be much smaller.

The reactions

$$O(^1D) + O^- \rightarrow O(^3P) + O(^3P) + e \tag{20a}$$

and

$$O(^1D) + O^- \rightarrow O(^3P) + O^{-\prime} \tag{20b}$$

---

* Neglecting vibrational excitation, (15) has an energy excess of 3·19 eV ($Y = a\,^1\Delta_g$) or 2·54 eV ($Y = b\,^1\Sigma_g^+$). For (16) very close resonance can occur; for example

$$O(^1D) + O_2(X\,^3\Sigma, v = 0) \rightarrow O(^3P_2) + O_2(b\,^1\Sigma_g^+, v = 2)$$

has an energy defect of only 0·004 eV.

† The reaction $O(^1D) + N_2 \rightarrow NO + N$ has an energy defect of 1·31 eV (or 0·12 eV using the HERZBERG dissociation energy), and will therefore not be important.

may give very close resonance,* but their importance is limited by the rate of production of O^-.

Reactions similar to (15)–(20) can occur with $N\ ^2D$ in place of $O\ ^1D$, and essentially similar remarks will apply, but much smaller coefficients will be of importance due to the great radiative lifetime of $N\ ^2D$. In addition there is

$$N(^2D) + O(^3P) \rightarrow N(^4S) + O(^1D) \tag{21}$$

with an energy excess of 0·41 eV. The reaction

$$N(^2P) + N_2 \rightarrow N(^2D) + N_2' \tag{22}$$

may be important in deactivating $N\ ^2P$.

3. THE O I FORBIDDEN LINES

3.1. *Observed relative intensities.* Dr. W. PETRIE (private communication) gives the intensity ratio $I(6300 + 64)/I(5577)$ as 0·40 for the region between the base and the level of maximum luminosity of steady homogeneous arcs (100–110 Km). He states that the value of this ratio remains constant for one minute or more. Since deactivation certainly occurs (cf. section 2.3) the lifetimes of the excited states will be considerably less than one minute and it therefore appears reasonable to assume that near equilibrium conditions are established.

At the upper border, 125–135 Km, $I(6300 + 64)/I(5577)$ varies between 1·0 and 1·4 (VEGARD, 1938). We adopt the value 1·2. MEINEL (1952) has published some very interesting results on the variation with angular elevation of the intensities of various spectral features. The following points may be noted:

(i) The N_2^+ negative group, the N_2^+ infra-red bands, the N_2 first positive group and the O I permitted lines, λ 7774 and 8446, all have nearly the same height-luminosity variation, which we will refer to as the main luminosity curve.

(ii) With decreasing height from the upper border the intensity of the red [O I] lines λ 6300 + 64, increases much less rapidly than does the main luminosity and has a broad maximum at a height slightly greater than that of the main maximum. This suggests that deactivation becomes increasingly important with decreasing height. However, below the maximum the intensity of the red line decreases less rapidly with decreasing altitude than does the main luminosity. Since deactivation is unlikely to become less important below the maximum, this behaviour suggests that an additional excitation mechanism becomes important in this region.

(iii) Above the level of maximum luminosity the green [O I] line, λ 5577, follows closely the main luminosity curve, but below this level the green line intensity decreases less rapidly with altitude, its variation being closely parallel to that of the red line. This behaviour is in agreement with the suggestion that an additional excitation mechanism for the [O I] lines becomes important at and below the maximum (100–110 Km).

* $O(^1D)$, with an excitation energy of 1·96 eV, may have just sufficient energy to detach an electron from O^- (electron affinity $\simeq 2{\cdot}2 \pm 0{\cdot}2$ eV), giving reaction (20a). The reaction (20b) assumes the existence of an excited state of O^- with near zero binding energy. The evidence for the existence of this state is by no means conclusive.

VOL. 4 1953

3.2. *Electron excitation and deactivation with a Maxwell distribution of electron velocities.* This has been discussed by BARBIER (1948). For a given intensity ratio the equilibrium equations may be solved for $n(e)$ as a function of $T(e)$. Table 2 gives the solutions obtained for the lower border, using both the exact equations given by BARBIER and the approximate equations of section 2.3. It will be seen that no serious error arises from the use of the approximate equations.

Table 2. Maxwell distribution of electron velocities

$I(6300 + 64)/I(5577) = 0{\cdot}40$

$T(e)$ °K	*Exact equations*			*Approximate equations*		
	$n(e)$	$n(^1D)/n(^3P)$	$P(5577)/n(^3P)$	$n(e)$	$n(^1D)/n(^3P)$	$P(5577)/n(^3P)$
15,670	$1{\cdot}0.10^8$	0·13	0·0030	$1{\cdot}1.10^8$	0·16	0·0033
31,340	$2{\cdot}8.10^7$	0·27	0·0063	$3{\cdot}6.10^7$	0·28	0·0057
78,400	$1{\cdot}6.10^7$	0·43	0·0098	$2{\cdot}3.10^7$	0·45	0·0090
156,700	$1{\cdot}5.10^7$	0·50	0·0114	$2{\cdot}3.10^7$	0·52	0·0105

No solution exists for $n(e)$ unless $T(e)$ is greater than 10^4 °K.

To account for the observed relative intensities it is seen that both $n(e)$ and $T(e)$ must be large, which leads to a high degree of excitation ($n(O, {}^1D)/n(O, {}^3P)$) and a high photon emission rate per oxygen atom ($P(5577)/n(O, {}^3P)$). Taking $n(O)$ at 105 Km to be $4{\cdot}3.10^{12}$ cm^{-3} (BATES and WITHERSPOON, 1952) and $t = 1$, Table 2 gives $\pi\,(5577) = 1{\cdot}3 - 4{\cdot}9.10^{15}$ cm^{-2} sec^{-1}. This is close to the extreme upper limit of paper 1 Table 2, and is certainly much brighter than the average display.

We may conclude that the assumptions made, that only electron collisions are important and that the electrons have a Maxwell distribution, must be discarded as untenable.

The assumption of a Maxwell distribution has previously been criticized by BATES (1949) from a consideration of the processes determining the distribution function. He divided the electrons into an active group, with density $n(eA)$, and a passive group, with density $n(eP)$, according to whether or not they were capable of exciting and ionizing the atmospheric constituents. Equating the number entering the passive group to the number leaving, he obtained an equation which reduces to

$$\zeta n(eA)n(\text{Tot.}) = \alpha_{\text{rec.}}\, n(eP)^2 \qquad (23)$$

when certain small terms are neglected. Here ζ is the coefficient associated with the process of degradation from the active group (taken to be 10^{-7} cm^3 sec^{-1}), $\alpha_{\text{rec.}}$ is the recombination coefficient (taken to be 10^{-8} cm^3 sec^{-1}) and n(Tot.) is the total particle density. From this it readily follows that $n(eA)$ is much less than $n(eP)$. We may put $n(eP) = n(e)$, the total electron density, without significant error.

It is convenient to introduce two effective temperatures, $T(eP)$ for the passive group and $T(eA)$ for the active group. The value of $T(eP)$ will be equal to or greater than the gas temperature. Since we are only interested in the deactivating effects of the passive electrons the exact value of $T(eP)$ is not important. We adopt the value of 10^3 °K, giving for electron deactivation of O I, $d_2 = 1{\cdot}62.10^{-9}n(e)$. Although $T(eA)$ has little physical significance, it is a convenient parameter for estimating the relative rates of various excitation processes.

3.3. *Deactivation mechanisms.* Using the observed intensity ratios and equation (8), we may determine the values of the deactivation probability d_2 as a function of the excitation ratio, S_2/S_3. The results of this calculation are given in Table 3, together with the electron density required if electron deactivation is assumed to be the sole deactivation mechanism.

Table 3. O I deactivation probability as a function of excitation ratio

S_2/S_3	*Lower border* $I(6300 + 64)/I(5577) = 0{\cdot}40$		*Upper border* $I(6300 + 64)/I(5577) = 1{\cdot}2$	
	$d_2\ sec^{-1}$	$n(e)\ cm^{-3}$	$d_2\ sec^{-1}$	$n(e)\ cm^{-3}$
0	0·011	$6{\cdot}8.10^6$	—	—
1	0·032	$2{\cdot}0.10^7$	0·0047	$2{\cdot}9.10^6$
5	0·12	$7{\cdot}2.10^7$	0·033	$2{\cdot}0.10^7$
12·5	0·28	$1{\cdot}7.10^8$	0·086	$5{\cdot}3.10^7$

For electron excitation the excitation ratio is given approximately by

$$\frac{S_2}{S_3} = 5{\cdot}5 \exp\left[(E_3 - E_2)/kT(eA)\right] \qquad (24)$$

Since the initial energy of the ejected electrons will be considerably greater than the energy required for excitation of the [O I] lines, it is probable that the number of electrons capable of exciting $O\,^1D$ is not much greater than the number capable of exciting $O\,^1S$. We might therefore expect the exponential in (24) to be not much greater than unity. The possibility of a larger value cannot, however, be excluded;* $S_2/S_3 = 12{\cdot}5$ corresponds to $kT(eA) = 2{\cdot}7$ eV. For excitation by dissociative recombination we may assume $\mathcal{S}_2/S_3 = 5$ (section 2.5).

In paper 1 it was concluded that electron densities between 10^7 and $10^8/cm^3$ occur in bright high latitude aurorae. Comparison with Table 3 shows that electron deactivation may well be the dominant deactivation mechanism. In order to make a more final decision on the importance of this process it would be of great interest to study the variation of the intensity ratio with the brightness of the display and with the electron density determined from radar soundings.

It may be noted that the values for d_2 obtained in Table 3 are consistent with

* If deactivation of $O\,^1D$ by electron impact is important this will have some tendency to lower $kT(eA)$.

VOL. 4 1953

the neglect of d_3 compared with A_{31} and A_{32} (assuming d_3 to be of a magnitude comparable to d_2). There does not appear to be any observational evidence which suggests that d_3 is not negligible.

There is some evidence for deactivation in the quiescent night sky,* when electron deactivation will certainly not occur. It is therefore of particular interest to see if deactivation by collisions with gas molecules could account for the auroral intensities. Using the particle densities of BATES and WITHERSPOON (1952) we consider the possibility of such collisions giving $d_2 = 0{\cdot}12$ at 105 Km and 0·033 at 130 Km.

Deactivation by O_2 at 105 Km would require $\gamma_{(16)}$ to be $6{\cdot}5.10^{-3}$. This is more than 10^3 times the VEGARD and KVIFTE value, but in view of the discussion of section 2.8 it can hardly be dismissed as impossible. However, deactivation by O_2 would decrease very rapidly with altitude and it cannot, therefore, be the sole mechanism of importance.

Deactivation by N_2 would require $\gamma_{(17)}$ to be 5.10^{-5} at 105 Km and $1{\cdot}4.10^{-4}$ at 130 Km. In view of the highly forbidden nature of (17) these probabilities appear to be rather large, but again it is not easy to dismiss them as impossible.

Deactivation by O^- is of particular interest since there is some evidence (STØRMER, 1937; VEGARD, 1936), that the oxygen red lines are enhanced in sunlit aurorae, for which the O^- concentration would be greatly reduced. An upper limit for the number of deactivations which can occur through the reactions (20) is given by the rate of production of O^- through the reaction

$$O(^3P) + e \rightarrow O^- + h\nu \tag{25}$$

The latter rate is $\eta n(O, {}^3P)n(e)$, where the most probable value for η is, according to BATES and MASSEY (1947), $1{\cdot}1.10^{-15}$ cm^3 sec^{-1}. Since the upper limit for the deactivation rate is proportional to $n(e)$, we consider the condition for deactivation by O^- to be more rapid than electron deactivation. This condition is

$$\eta n(O, {}^3P) > \alpha^e_{21} n(O, {}^1D), \tag{26}$$

which limits the value of $n(O, {}^1D)$, and hence the total brightness. The number of red line quanta per cm^2 column per sec is $\pi(6300 + 64) = 10^5 t A_{21} n(O, {}^1D)$. Denoting the ratio $I(6300 + 64)/I(5577)$ by r, and putting $\eta = 1{\cdot}1.10^{-15}$ cm^3 sec^{-1} and $\alpha^e_{21} = 1{\cdot}62.10^{-9}$ cm^3 sec^{-1}, the condition (26) gives

$$\pi(5577) < 5.10^{-4} n(O, {}^3P)t/r \tag{27}$$

For $t = 1$ this gives $\pi\,(5577) < 5.10^9$ at 105 Km and $< 2.10^8$ at 130 Km. Comparison with Table 2 of paper 1 shows that O^- deactivation could only be important at 105 Km for the faintest aurorae. For 130 Km the upper limit is comparable to the night sky brightness. We may conclude that deactivation by O^- will in general be less important than electron deactivation, but it is just possible that removal of O^- in sunlight could be responsible for some enhancement of the red lines for the faintest aurorae.

* This will be further discussed in a later paper.

3.4. *Excitation mechanisms.* We consider first the intensity resulting from electron excitation. The most direct approach is provided by considering the rate of production of active electrons, $q(eA)$ and the processes responsible for their degradation. We assume that one active electron is produced for each N_2^+ N.G. quantum emitted, and since other ionization processes will also contribute, $q(eA)$ must be at least as great as the emission rate of N.G. quanta, P(N.G.). It was shown in paper 1 section 4.2 that $P(\text{N.G.}) \cong 10P(5577)$, giving $q(eA) \geqslant 10P(5577)$.

The cross sections being of order gas kinetic, electron excitation of O I will be one of the major mechanisms responsible for the degradation of the active electrons. Although other processes will compete, the most important being electron excitation* of N_2, each active electron will have initially sufficient energy to excite O I forbidden lines several times. We may therefore estimate that the number of O I forbidden line excitations per active electron will be of order unity, and that the total O I excitation rate due to electron impact will be of order 10 $P(5577)$ or greater.

It is seen that the estimated excitation rate is greater than the observed emission rate. This is to be expected, due to deactivation effects. With electron excitation the excitation rate is equal to or greater than 6·5 P(5577) (cf. (24)). It may be concluded that electron excitation is in general sufficient to produce the observed [O I] emissions.

Above the level where oxygen is mainly in the atomic form it is difficult to see what mechanism other than electron excitation could be operative. However, it has already been pointed out in section 3.1 that there is evidence for an additional mechanism becoming important at the lower border (90–100 Km (BERKEY and GARTLEIN, 1952)), which is the region where the molecular oxygen density becomes large. We consider the possibility of identifying this mechanism with the dissociative recombination reaction (12).

This reaction gives two oxygen atoms per recombination, and assuming the average distribution of these among the ground configuration terms to be in the ratio of the statistical weights, we obtain for the photon emission rate in the green line due to this mechanism

$$P(5577 \mid O_2^+) = \frac{2}{15} q(O_2^+) \tag{28}$$

Although O_2^+ bands have been identified in the auroral spectrum (PETRIE and SMALL, 1952a; GARTLEIN and SHERMAN, 1952), no intensity estimates from which the O_2 ionization rate, $q(O_2^+)$, could be estimated appear to have been published. We may make a provisional estimate by assuming that O_2 and N_2 have equal ionization probabilities, giving

$$q(O_2^+) = q(N_2^+)n(O_2)/n(N_2) \tag{29}$$

Taking $q(N_2^+) \geqslant P(\text{N.G.})$ this gives

$$\frac{P(5577 \mid O_2^+)}{P(\text{N.G.})} \geqslant \frac{2}{15} \frac{n(O_2)}{n(N_2)} \tag{30}$$

* The N_2 2nd P.G. is probably excited mainly by electron impact (PETRIE and SMALL, 1952b).

VOL. 4 1953

The observed ratio may be taken as before to be $P(5577 \mid \text{Obs.})/P(\text{N.G.}) = 0{\cdot}1$. We obtain finally

$$\frac{P(5577 \mid O_2^+)}{P(5577 \mid \text{Obs.})} \geqslant \frac{20}{15}\frac{n(O_2)}{n(N_2)}$$

The densities of BATES and WITHERSPOON (1952) give the following values:

Table 4

Height Km	$\frac{20}{15}\frac{n(O_2)}{n(N_2)}$
100	0.046
95	0.28
90	0.33

It is seen that dissociative recombination appears to be capable of accounting for at least a third of the observed emission at the lower border, and may well account for practically the whole emission. Its effectiveness will, however, decrease very rapidly above the lower border.

In discussing the excitation and deactivation mechanisms we have assumed the red and green lines to originate from the same regions in space. This may not be strictly true, since the highly excited regions responsible for the main luminosity may be embedded in more extensive regions containing a lower density of moderate energy electrons. Such regions would give preferential excitation of the red lines due to the larger electron excitation value of S_2/S_3 and to the smaller electron deactivation probability. To make further progress it would be of great interest to observe the intensity variations in horizontal cross sections across auroral forms. MEINEL (1952) states that the red line intensity is not well correlated with auroral forms but is smeared out over large regions of the sky. While this may well be correct it is not clear whether the observations of MEINEL establish it. Apparently he considered only the intensity variation in vertical cross sections, which may be largely determined by the altitude variation of excitation and deactivation mechanisms.

It would also be of great interest to observe the variation with time of the red to green intensity ratio. Here again we might expect that the red lines would increase relative to the green in the later stages of a display due to an increase of S_2/S_3 and a decrease in electron deactivation.

4. THE N I FORBIDDEN LINES

4.1. *Observed relative intensities.* The relative intensities of the [N I] lines are much less certain than those of the [O I] lines. This is partly a consequence of the fact that λ 5199 is a faint feature and that λ 3467 is in a region where the ultraviolet correction factor is large. Additional uncertainties arise from the fact that the intensities vary, relative to the main luminosity, both with altitude and with time.

VEGARD and KVIFTE (1945) give the uncorrected intensity $I_O(3467)$ to be 3·0; using the correction factor f_λ we obtain a corrected intensity* $I_C(3467) = 120$. The variation of $I(3467)$ with altitude has been studied by BERNARD (1941) who found $I_O(3467)$ to vary by a factor of 10, relative to $I_O(3914)$, between the base and the summit. It was also found that the forbidden VEGARD-KAPLAN bands (N_2, $A\,^3\Sigma \rightarrow X\,^1\Sigma$) varied in a very similar way. While some variation might arise from incomplete allowance for atmospheric absorption, there can be little doubt that these results indicate a real variation in the emission rate. The most obvious interpretation is that there is appreciable deactivation at the lower altitudes for both λ 3467 and the VEGARD-KAPLAN bands.

GARTLEIN and SHERMAN (1952) report that λ 5199 appears rather late in the display, when nitrogen bands and N II lines have been strong, that it persists as the rest of the spectrum dies away, and that $I(5199)$ rarely exceeds $I(4708)$ (N_2^+ N.G.; 0,2). Using the corrected VEGARD and KVIFTE intensities we estimate that $I(5199)$ rarely exceeds 14. The variability of $I(5199)$ is confirmed by Dr. D. M. HUNTEN (private communication) who states that, while $I(5199)$ is comparable to $I(5228)$ (N_2^+ N.G.; 0,3) on photographic plates with exposures of an hour or two, it has not been detected using a high-speed photo-multiplier ($I(5199)$ probably less than one tenth of $I(5288)$). Making some allowance for the variability of this line, Dr. W. PETRIE (private communication) estimates that $I(5199) = 1{\cdot}5$.

In discussing the theoretical interpretation we will use both the corrected and uncorrected VEGARD and KVIFTE values for $I(3467)$ and PETRIE's estimate for $I(5199)$.

4.2. *Deactivation mechanisms.* In Table 5 we give d_2 as a function of S_2/S_3 and the corresponding values of $n(e)$ for electron deactivation, using both the corrected and uncorrected VEGARD and KVIFTE values for $I(3467)$ and neglecting

Table 5. N I deactivation probability (d_2) as a function of excitation ratio

S_2/S_3	*Without u-v correction* $I(5199)/I(3467) = 0{\cdot}5$		*With u-v correction* $I(5199)/I(3467) = 0{\cdot}0125$	
	d_2 *sec*$^{-1}$	$n(e)$ *cm*$^{-3}$	d_2 *sec*$^{-1}$	$n(e)$ *cm*$^{-3}$
0	0·00019	$2{\cdot}2.10^5$	0·0079	$9{\cdot}3.10^6$
1	0·00039	$4{\cdot}6.10^5$	0·016	$1{\cdot}9.10^7$
10/6	0·00053	$6{\cdot}3.10^5$	0·022	$2{\cdot}2.10^7$
2·56	0·00072	$8{\cdot}6.10^5$	0·029	$3{\cdot}5.10^7$

d_3. For similar mechanisms the S_2/S_3 ratios to be expected for N I are smaller than those for O I; the value 10/6 corresponds to the statistical weight ratio and 2·56 corresponds to electron excitation with $kT(eA) = 2{\cdot}7$ eV.

* With this value for $I(3467)$, relative to $I(5577) = 100$, the infra-red lines of [N I] at λ 10407, 10398 would have an intensity of nearly 600.

VOL. 4 1953

Without the ultra-violet correction factor the values of d_2 for N I are much less than the values for O I given in Table 3, and the calculated electron densities, which represent upper limits, are seen to be improbably low (paper 1, section 3). If these values of $n(e)$ were accepted it would be necessary to postulate a deactivation mechanism for O I, other than electron deactivation, which was not effective for N I. This again appears to be improbable. In general it appears to be very difficult to account for the observed intensities if the ultra-violet correction is not used.* The only explanation which would appear at all plausible would be to assume that a reaction such as (22) gives a very large value for d_{32}, which would be equivalent to adopting a value of S_2/S_3 much greater than the values in Table 5. It would, however, be necessary to postulate an unexpectedly large cross section for (22), and in addition it would not be at all clear why the intensities of the Vegard-Kaplan bands should vary in a similar way to $I(3467)$.

With the correction factor the values of d_2, and the calculated electron densities, are in reasonable agreement with those obtained for O I. With these densities the neglect of d_3 and d_{32} is not entirely justified. In Table 6 we give $n(e)$ for electron deactivation, including the d_3 and d_{32} terms. The values obtained are in close agreement with the values for the lower border given in Table 3 for electron deactivation of O I.

Table 6. N I electron deactivation with d_3 and d_{32} terms (allowing for u-v correction)

S_2/S_3	$n(e)$ cm^{-3}
0	$1{\cdot}2.10^7$
1	$3{\cdot}6.10^7$
10/6	$7{\cdot}1.10^7$
2·56	$2{\cdot}3.10^8$

A reduction in $I(3467)$ by a factor of 10 in going from the base to the summit, suggested by the observations of BERNARD, could be obtained if $n(e)$ was $3{\cdot}5.10^8/cm^3$ at the base and $1.10^7/cm^3$ at the summit. Values of this order appear to be possible for a bright display.

It may be concluded that electron deactivation appears to be capable of accounting for the main features of the relative intensities.

4.3. *Excitation mechanisms.* In discussing excitation mechanisms we will use only the corrected intensity for λ 3467.

We consider first the possibility that both the [O I] and [N I] lines are excited and deactivated by electron impact. The equations of equilibrium give

$$\frac{I(3467)}{I(5577)} = \frac{5577}{3467} \cdot \frac{\left(\dfrac{A_{31}\alpha^e_{13}}{A_{31} + A_{32} + (\alpha^e_{31} + \alpha^e_{32})n(e)}\right)_{\text{N I}}}{\left(\dfrac{A_{32}\alpha^e_{13}}{A_{31} + A_{32} + (\alpha^e_{31} + \alpha^e_{32})n(e)}\right)_{\text{O I}}} \cdot \frac{n(N)}{n(O)} \qquad (31)$$

* It was this problem which first led the author to examine the possibility of a systematic error in the ultra-violet intensities.

Substituting numerical values we obtain

$$\frac{I(3467)}{I(5577)} = 0{\cdot}77 \cdot \frac{1 + 1{\cdot}3.10^{-9}n(e)}{1 + 2{\cdot}5.10^{-8}n(e)} \cdot \frac{n(N)}{n(O)} \tag{32}$$

where we have assumed that the number of electrons capable of exciting N I $^2P(E_3 = 3{\cdot}56$ eV) does not differ significantly from the number capable of exciting O I $^1S(E_3 = 4{\cdot}17$ eV). Table 7 gives calculated values of $n(N)/n(O)$ obtained with $I(3467)/I(5577) = 120/100$. Very similar results may be obtained on putting $I(5199)/I(6300 + 64) = 1{\cdot}5/40$.

Table 7. $n(N)/n(O)$ *for electron excitation*

$n(e)$ cm^{-3}	$n(N)/n(O)$
10^6	1·6
10^7	2·0
10^8	5·5

Assuming the chemical composition of the atmosphere at auroral levels to be the same as at ground level, and that nearly all the oxygen is dissociated (BATES and WITHERSPOON, 1952), complete nitrogen dissociation would correspond to $n(N)/n(O) = 4{\cdot}0$. It is seen that a very high degree of nitrogen dissociation would be required if the [N I] lines were excited by electron impact. This appears to be improbable on theoretical ground (BATES, 1954), and if most of the nitrogen were in the atomic form it would be difficult to account for the great intensities of molecular nitrogen bands in the auroral spectrum.

The other obvious possibility is that dissociative recombination is responsible for exciting the [N I] lines. Assuming as in 3.4 that $q(N_2^+) \geqslant P$(N.G.), that P(N.G.) $\simeq 10.P(5577)$ and that the atoms produced by (13) are distributed among the ground configuration terms in the ratio of the statistical weights, we obtain S_3(N I) $\geqslant 6.P(5577)$. The calculated lower limits for $I(3467)$ are given in Table 8, allowing for electron deactivation.

Table 8. $I(3467)$ *due to dissociative recombination*

$n(e)$ cm^{-3}	*Lower limit for* $I(3467)$ $(I(5577) = 100)$
10^6	62
10^7	50
10^8	18

Recalling that $q(N_2^+)$ will be somewhat greater than P(N.G), this mechanism appears to be capable of accounting for the observed intensity, $I(3467) = 120$.

VOL. 4 1953

4.4. *Variations throughout the display.* As mentioned in section 4.1, λ 5199 appears to be enhanced relative to other features in the later stages of a display. It is readily shown to be very unlikely that dissociative recombination could give a sufficient increase in $n(N)$ to influence the importance of electron excitation. It is much more likely that the relative enhancement of λ 5199 is a consequence of electron deactivation. We would expect the intensity of any feature for which electron deactivation is important to vary as $n(e)$, whereas without deactivation we would expect the intensity to vary as $n(e)^2$. This difference could account for a relative enhancement of λ 5199 as $n(e)$ decays.

5. Forbidden O II and N II Lines

5.1. *The [O II] lines.* Several workers (Vegard and Kvifte, 1945; Barbier and Williams, 1950) have suggested that a faint auroral line at 3727 Å could be identified with the forbidden $^2D \rightarrow {}^4S$ transition of [O II], which has the very low transition probability of $A = 7{\cdot}7.10^{-5}$ sec^{-1} (Garstang, 1952). Petrie (1952) has recently criticized this identification, since he finds that the O[II] $^2P \rightarrow {}^2D$ lines at λ 7319, 7330 Å are definitely missing despite their much larger A value of 0.135 sec^{-1} (Pasternack, 1940; Garstang, 1952). We may discuss this further by considering the excitation rate, S_2(O II) which would be required. Vegard and Kvifte (1945) give $I_O(3727) = 1{\cdot}0$, corresponding to $I_C(3727) = 20$. The most instructive procedure is to compare S_2(O II) with S_2(O I), taking $I(6300 + 64) = 40$ and assuming electron deactivation (Seaton, 1953b). Table 9 shows that this identification would require S_2(O II) to be several hundred times larger than S_2(O I), which appears extremely improbable, and confirms Petrie's view that the feature at λ 3727 is not due to [O II].

Table 9

$n(e)$ cm^{-3}	S_2(O II)/S_2(O I) *required if* λ 3727 *is identified with* [O II] $^2D \rightarrow {}^4S$
10^6	130
10^7	550
10^8	810

5.2. *The [N II] lines.* Petrie (1952) reports a sharp atomic feature at λ 5755 and points out that the identification of this with [N II] $^1S \rightarrow {}^1D$ ($A = 1{\cdot}08$ sec^{-1}) would be consistent with the absence of the [N II] $^1D \rightarrow {}^3P$ lines at λ 6548, 6584, which have the much lower A-value of 0·0040 sec^{-1} (Garstang, 1951). He suggests that N II ^1S may be populated by simultaneous ionization and excitation* of N I ^2D giving N II ^1S and other N II singlet states which will

* Petrie remarks that simultaneous ionization and excitation of the N I ^4S ground state cannot give N II singlet states. This is not true for ionization by electron impact, for which the singlet states can be formed without violation of the spin conservation rule.

cascade to ^1S. This mechanism does not appear very convincing, since N I ^2D is a comparatively rare constituent, its density being unlikely to exceed 10^{-4} times the total particle density. A more promising mechanism is provided by (2), dissociative ionization of N_2.

Dissociative ionization would be expected to produce excited N I and excited N II at comparable rates. Since the excitation rates required for the forbidden [N I] lines are much greater than the rates required for the [N II] lines, it is considered improbable that (2) is important as an N I excitation mechanism.

5.3. *The [O III] lines.* MEINEL (1951) has pointed out that identification of a feature at λ 5007 with the [O III] nebular line $^1D \rightarrow {}^3P_2$ must be rejected, since the other member of the doublet, λ 4959 $^1D \rightarrow {}^3P_1$ has never been observed.

6. CONCLUSIONS

We summarize briefly the processes which are considered most likely to be of importance:

(i) At and above the level of maximum luminosity the [O I] lines are excited by electron impact.

(ii) Below the level of maximum luminosity dissociative recombination of O_2^+ contributes to excitation of the [O I] lines.

(iii) The [N I] lines are excited by dissociative recombination of N_2^+; electron excitation is less likely to be of importance.

(iv) Both O I and N I metastable states are deactivated by electron impact. Deactivation resulting from collisions with gas molecules is less likely to be of importance.

(v) The only other forbidden atomic line for which the identification is not in doubt is [N II] $^1S \rightarrow {}^1D$ at λ 5755. This could be excited by dissociative ionization of N_2.

Acknowledgments—The author is indebted to Professor H. S. W. MASSEY, F.R.S., and the Professor D. R. BATES for numerous valuable suggestions and comments. He is also indebted to Dr. W. PETRIE and to Dr. D. M. HUNTEN for their generosity in supplying details of their observational results in advance of publication.

REFERENCES

BARBIER, D.	1948	Ann. d'Astrophys. **11**, 193
BARBIER, D. and WILLIAMS, D. R.	1950	J. Geophys. Res. **55**, 401
BATES, D. R.	1949	Proc. Roy. Soc. A **196**, 217
BATES, D. R. and MASSEY, H. S. W.	1947	Proc. Roy. Soc. A **192**, 1
BATES, D. R., MASSEY, H. S. W., and PEARSE, R. W. B.	1948	Report Gassiot Ctte., Phys. Soc. Lond., p. 97
BATES, D. R.	1954	The Physics of the Upper Atmosphere, Ch. 12 of The Solar System (Ed. G. P. KUIPER), Vol. 2, University of Chicago Press
BATES, D. R. and WITHERSPOON, A. E.	1952	Mon. Not. R.A.S. **112**, 101
BERKEY, D. K. and GARTLEIN, C. W.	1952	Mém. Soc. Roy. Liège **12**, 199
BERNARD, R.	1941	Ann. d'Astrophys. **4**, 31
DOUGLAS, A. E.	1952	Can. J. Phys. **30**, 302

VOL. 4 1953

GARSTANG, R. H.	1951	Mon. Not. R.A.S. **111,** 115
	1952	Astrophys. J. **115,** 506
GARTLEIN, C. W. and SHERMAN, D. F.	1952	Mém. Soc. Roy. Liège **12,** 187
GAYDON, A. G.	1953	Dissociation Energies and Spectra of Diatomic Molecules, 2nd ed., Chapman and Hall, London, Ch. IX
GHOSH, S. N.	1946	Proc. Nat. Inst. Sci. India **12,** 405
HERZBERG, G.	1950	Spectra of Diatomic Molecules, 2nd ed., Van Nostrand, New York, p. 448
MASSEY, H. S. W.	1952	Phil. Mag. Suppl. **1,** 395
MASSEY, H. S. W. and BURHOP, E. H. S.	1952	Electronic and Ionic Impact Phenomena. Oxford, Clarendon Press, p. 263
MEINEL, A. B.	1951	Rep. Progr. Phys. **14,** 121
	1952	Mém. Soc. Roy. Liège **12,** 203
MEINEL, A. B. and FAN, C. Y.	1952	Astrophys. J. **115,** 330
MITRA, S. K.	1946	Nature **157,** 692
	1951	Nature **167,** 897
NICOLET, M.	1945	Ciel et Terre **61,** 287
OLDENBERG, O.	1952	Phys. Rev. **87,** 786
PASTERNACK, S.	1940	Astrophys. J. **92,** 129
PETRIE, W.	1952	Phys. Rev. **87,** 1002
PETRIE, W. and SMALL, R.	1952a	Astrophys. J. **116,** 433
	1952b	J. Geophys. Res. **57,** 51
SAYERS, N. D. and EMELEUS, K. G.	1952	Proc. Phys. Soc. A **65,** 219
SEATON, M. J.	1953a	Phil. Trans. A **245,** 469
	1953b	Proc. Roy. Soc. A **218,** 400
	1954a	J. Atmosph. Terr. Phys. **4,** 285
	1954b	Mon. Not. R.A.S. in course of publication
STØRMER, C.	1937	Nature **139,** 584
SWINGS, P.	1948	The Atmospheres of the Earth and Planets (ed. G. P. KUIPER), University of Chicago Press, p. 159
VEGARD, L.	1938	Nature **141,** 200
	1936	Nature **138,** 930
VEGARD, L. and KVIFTE, G.	1945	Geofys. Publ. **16,** No. 7
	1947	Geofys. Publ. **17,** No. 1

Journal of Atmospheric and Terrestrial Physics, 1954, Vol. 4, pp. 314 to 321. Pergamon Press Ltd., London

The variation of electrical conductivity of air within sealed rooms

L. G. Smith and G. F. Schilling

Institute of Geophysics, University of California, Los Angeles 24, California

(*Received* 3 *October* 1953)

ABSTRACT

The electrical conductivity of air within sealed rooms has been found to increase regularly, with some fluctuations superimposed, towards a limiting value. This variation is compared with Schweidler's theory of ion-equilibrium and the increase in conductivity attributed to a decrease in Z, the total concentration of nuclei. It is deduced that the rate of change of Z can best be represented by

$$\frac{dZ}{dt} = -\psi Z^{3/2}$$

where ψ is a constant with the value $2{\cdot}5 \times 10^{-7}$ cm$^{3/2}$ sec^{-1}. This equation is shown to be close to that obtained by Nolan from direct observations of nuclei.

1. Introduction

The observations of the electrical conductivity of air within sealed rooms which are discussed in this paper were obtained incidentally in the course of an investigation of atmospheric electrical parameters. Though not representing an exhaustive study they were felt to be of considerable interest as showing a new approach to investigations of ion-equilibrium in the atmosphere and the possibility of extension to other problems.

Results of three experiments will be discussed. As will be shown, however, the first one was not carried to its logical conclusion and has to be regarded as incomplete.

2. Experimental Details

The conductivity meter used in these experiments is of the type originated by Gerdien, with the electrical system modified to give a continuous record of atmospheric conductivity. The calibration of the instrument is known accurately. The value of the polar conductivity will be quoted in reciprocal ohm-meters times 10^{-15} and called "units" except when otherwise stated.*

The first two experiments were performed in a room on the second floor of a building on the campus of the University of California in West Los Angeles. The approximate dimensions of the room are: 9 m long, 6 m wide, and 4 m high. The room is normally used as a workshop and contains much equipment. A third experiment was performed in an adjacent room, the dimensions of which are: 4 m long, 2 m wide, and 4 m high. This room had not been occupied at any time since construction of the building and was free from equipment, except for a table and bench. The ventilation system was switched off and the vents and the door sealed for the periods of the experiments.

3. Observations

The variation of the positive conductivity with time during the three experiments is shown in Fig. 1. In the case of the first experiment the room was unsealed

* (e.s.u. $\times 10^{-4}$) = ($\Omega^{-1}m^{-1} \times 10^{-15}$) $\times$ 0·09.

VOL. 4 1953

after a period of 18 hours 35 minutes. The increase was linear from an initial value of 3·1 units to a value of 36·7 units at the time of termination, a mean rate of 1·81 units per hour. The maximum deviation from the best linear relation is 0·7 units.

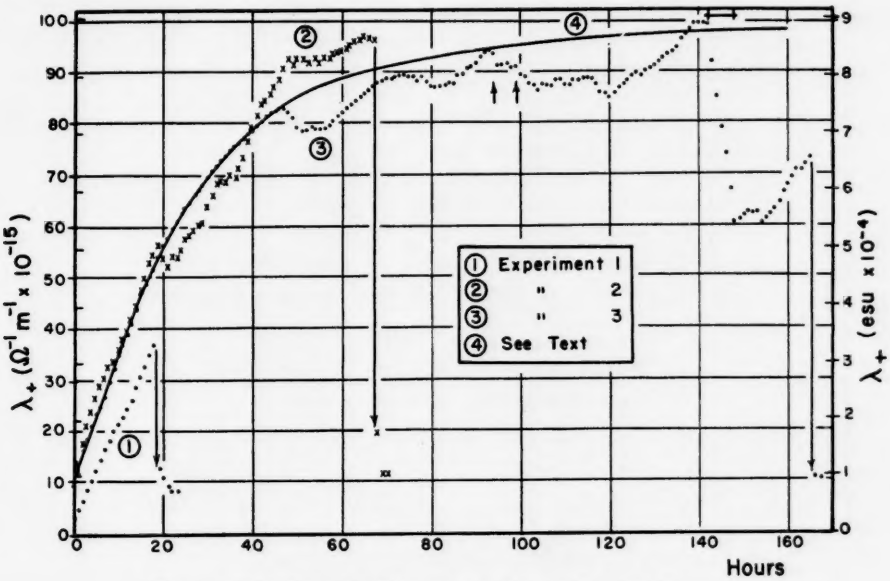


Fig. 1. Variation of positive conductivity within sealed rooms.

The record showed small, rapid fluctuations ("noise" in form of spikes of the duration of seconds) which increased in amplitude from about 0·5 to 1·0 units as the mean value of conductivity increased. When the room was unsealed, the conductivity decreased suddenly.

During the second experiment, the initially linear increase of conductivity changed later to an approximately assymptotic rise. A maximum value of 97·0 units had been reached after 68 hours when the experiment was terminated. It can be seen in the figure that irregularities occurred during the increase, but no change in the general character of the assymptotic approach is evident. The record again showed small, rapid fluctuations (noise) increasing steadily in amplitude from 0·4 units to 1·5 units. When the room was opened the conductivity fell within ten minutes and then continued the type of diurnal variation normally observed in the room.

The third experiment was performed in a much smaller room, but the figure shows that the general result is surprisingly similar to those of the previous experiments. The door was pounded 94 hours 45 minutes after the initial sealing of the room causing a fall of conductivity by 2·3 units. The door was pounded again 4 hours 35 minutes later producing an immediate decrease of 2·0 units (see short arrows in Fig. 1). After some irregularities during the subsequent hours the conductivity increased again to the approximate maximum value previously noted.

142 hours 40 minutes after the start of the experiment, the door was partly unsealed by removing a strip of tape 25 cm long. The conductivity began to decrease rapidly and continued so until the room was resealed 5 hours later.

The experiment was finally terminated after a total period of 165 hours. The rapid fluctuations (noise) did not show the regular increase in amplitude that was noted in the previous experiments, but showed frequently a small variation in amplitude with a period of about ten minutes.

4. Discussion

4.1. *General considerations*

The experiments showed that the conductivity of air in a completely sealed room increases (with some irregularities) as though towards a certain limiting value. The rate of increase decreases as this value is approached.

The rates of increase during the first twenty hours were as follows:

Experiment 1: 1·8 units per hour
Experiment 2: 2·4 units per hour
Experiment 3: 2·3 units per hour

Consistency of these values is such as to prohibit any close association with differences in the experimental conditions. Whereas experiments 1 and 2 were performed in a room of 216 m^3 volume, experiment 3 took place in a totally different room of 32 m^3 volume. Yet the initial rates of increase in experiments 2 and 3 are identical within possible instrumental errors.

VOL. 4 1953

All three records of conductivity show irregular variations in form of hour-to-hour fluctuations superimposed upon the general character of the curves. These fluctuations become larger as the conductivity approaches the limiting value. The amplitudes of the rapid fluctuations show a small increase during the earlier period but irregular behaviour when the limiting value is approached.

The importance of completely sealing the room was shown, in the third experiment, by the effect of partially unsealing the room. The conductivity fell at a rate which would have reduced it to the initial value within 15 hours. Since it has not been established that the rooms were sealed completely air tight, the possibility exists that some of the fluctuations were due to a partial exchange of air. However, it is conceivable that they represent inhomogeneities in the conductivity rather than changes in the mean value.

4.2. *The cause of the increase in conductivity*

It follows from the definition of polar conductivity, λ, and of the mobility of ions, k, that

$$\lambda_{\pm} = n_{\pm} e k_{\pm} \qquad (1)$$

where n is the concentration of ions and e is the average charge per ion, assumed to be equal to the electronic charge, a universal constant.

An increase in conductivity such as was observed in the present series of experiments could be interpreted, therefore, as being due to an increase in the concentration or in the mobility of the ions, or both. Since there is no evidence

that the mobility would change systematically, we shall attribute the change to variations in the concentration of ions and consider why this should increase towards a limiting value.

4.3. *The role of nuclei*

The equilibrium number of ions in a given space is determined by a balance between the production and removal of ions. In the free atmosphere small ions are produced by various radiations and are removed by two processes: (*a*) recombination and (*b*) combination with nuclei, both charged and uncharged. The balance equation for small ions of either sign is then:

$$q = \alpha n^2 + \eta N n + \eta_0 N_0 n \tag{2}$$

where q is the rate of production of ion-pairs, α is the recombination coefficient for small ions, n, N, and N_0 are the concentrations of small ions, of charged nuclei of either sign (large ions), and of uncharged nuclei, and η and η_0 are the combination coefficients between small ions and charged and uncharged nuclei respectively.

SCHWEIDLER (1918;1919) has shown, by considering a similar equilibrium for nuclei, that the above equation can be expressed in the form

$$q = \alpha n^2 + \beta Z n \tag{3}$$

where β is a quasi-combination coefficient defined by

$$\beta = \frac{2\eta_0 \eta}{(2\eta_0 + \eta)} \tag{3a}$$

and $Z = (N_0 + 2N)$ is the total concentration of nuclei.

Although equation (3) explains qualitatively the variation of conductivity observed in the atmosphere, it has been reported to be inadequate quantitatively. This may be due in part to the difficulties of making accurate measurements of the variables involved and to uncertainty in the value of β. In the absence of any definite evidence to the contrary we shall proceed on the assumption that it is correct.

In that case the increase in the concentration of small ions can be explained as an increase in the rate of production of ions or as a decrease in the concentration of nuclei (or both effects), neglecting, for the present, any possibility of variation in α and β. In the experiments in sealed rooms there is no reason to believe that the rate of production of ions increases systematically; we consider, therefore, the increase in conductivity to be due primarily to a decrease in the nuclei concentration.

4.4. *The limiting value of conductivity*

The ion concentration, n_∞, corresponding to elimination of all nuclei, is given by

$$q = \alpha n_\infty^2 \tag{4}$$

Thus, knowing the limiting value of conductivity, we can find n_∞ from equation (1) and hence compute the rate of production of ions. In practice the limiting value is not known exactly, but a lower limit for the value of q can be determined from

the maximum value of conductivity observed during the experiment. The maximum values were 97·0 units and 99·5 units in experiments two and three, respectively; whence, using the following constants:

$$\alpha = 1{\cdot}6 \times 10^{-6} \text{ cm}^3 \text{ sec}^{-1}$$
$$e = 1{\cdot}6 \times 10^{-19} \text{ coulomb}$$
$$k = 1{\cdot}4 \text{ cm per second per volt per cm}$$

we find the rates of production are, at least, 29 and 30 ion-pairs per cm^3 per second in the two cases.

The normal value in the atmosphere is considered to be about 10 ion-pairs per cm^3 per second. A test in which the conductivity apparatus was converted to an ionization meter by sealing the ends of the outer tube showed a rate of 30 ion-pairs per cm^3 per second in the laboratory room (not sealed). A further experiment in which a cage of wire gauze was used as the ionization chamber gave a value of 38 ion-pairs per cm^3 per second. It appears, therefore, that the rate of production of ions is unusually high in the rooms used for the experiments.

4.5. *The variation of nucleus concentration*

Substituting in equation (3) the value of q given by equation (4), and rearranging we have

$$\beta Z = \alpha \left\{ \frac{n_\infty^2 - n^2}{n} \right\} \tag{5}$$

The values of βZ have been computed for each hour of experiments two and three, and it has been found that the variation of Z during these experiments can be represented by

$$(\beta Z)^{-\frac{1}{2}} - (\beta Z)_0^{-\frac{1}{2}} = \frac{\phi t}{2} \tag{6}$$

where $(\beta Z)_0$ is the initial value of $\beta Z(t = 0)$ and ϕ is a constant having the value $1{\cdot}8 \times 10^{-4}$ sec$^{-\frac{1}{2}}$.

The variation of conductivity corresponding to this variation in βZ is entered as curve 4 in Fig. 1. It shows good agreement with the initial part of the third experiment and shows the tendency towards a limiting value.

The equation can also be represented in differential form by

$$\frac{\mathrm{d}(\beta Z)}{\mathrm{d}t} = -\phi(\beta Z)^{3/2} \tag{7}$$

4.6. *The value of β*

The equations derived above, representing the decrease in nucleus concentration in sealed rooms, cannot be compared with direct observations obtained by others until some value is given to β.

Values of η_0 and η, computed by Nolan and Kennan (1949) for nuclei of different radii, are given in Table 1 together with the value of β derived from them. It is seen that there is a roughly linear relation between β and the radius of the nuclei.

VOL. 4 1953

If the nuclei retain the same average size during the course of the experiment, β will remain constant. However, there may be a systematic change in the average size of the nuclei, for example by a selective diffusion process or by coagulation.

Table 1

Radius (*cm*)	η_0 ($cm^3\ sec^{-1}$)	η ($cm^3\ sec^{-1}$)	β ($cm^3\ sec^{-1}$)
$1{\cdot}0 \times 10^{-6}$	$0{\cdot}28 \times 10^{-6}$	$2{\cdot}7 \times 10^{-6}$	$0{\cdot}47 \times 10^{-6}$
2·0	1·0	3·4	1·3
3·0	2·0	4·4	2·1
4·0	3·2	5·6	2·9
5·0	4·6	7·0	4·0
6·0	5·9	8·3	4·9

In the latter case of coagulation (with no diffusion to the walls) the radius of the nuclei (and hence β) will be proportional to $Z^{-1/3}$.

Equation (7) becomes:
for β constant,

$$\frac{dZ}{dt} = -\psi Z^{3/2} \tag{7a}$$

for β proportional to $Z^{-1/3}$,

$$\frac{dZ}{dt} = -\psi' Z^{4/3} \tag{7b}$$

The difference between the exponents of Z in the two cases is relatively small although Z changes over a large range. It is concluded therefore that no serious error can be introduced into equation (7) by making the simplifying assumption that β is a constant during the experiment. We shall assume the value 2×10^{-6} $cm^3\ sec^{-1}$. This is the value in Table 1 corresponding to nuclei of radius 3×10^{-6} cm, found by Nolan and Doherty (1950) to be the average size of atmospheric nuclei.

Substituting this value for β in equation (7) we have

$$\frac{dZ}{dt} = -\psi Z^{3/2} \tag{7a}$$

where ψ is a constant with the value $2{\cdot}5 \times 10^{-7}$ $cm^{3/2}\ sec^{-1}$. Curve 4 then represents a decrease in concentration from 30,000 to 200 nuclei per cm^3.

Further observations were clearly needed to confirm certain assumptions which have been made in this section. Conductivity tests in a smaller room using a dust precipitator were found to be in general agreement with the explanation given above in terms of the reduction in concentration of nuclei. Actual measurements of large ions and nuclei concentrations which would have given a direct check were beyond the scope of the investigation.

5. CONCLUSIONS

The disappearance of nuclei in a closed vessel has been investigated by NOLAN (1941) and by NOLAN and KENNAN (1949). They find that the decrease in concentration of nuclei can be represented by

$$\frac{dZ}{dt} = -(\rho Z^2 + \sigma Z) \tag{8}$$

where ρ and σ are constants. The ρZ^2 term, which predominates at higher concentrations, they consider to represent loss of nuclei by coagulation and the σZ term, which is the more important at lower concentrations, to represent loss by diffusion to the walls. The experimentally determined values of ρ and σ are stated to be in close agreement with theoretical values.

The corresponding relation deduced from the observations made during the present investigation is

$$\frac{dZ}{dt} = -\psi Z^{3/2} \tag{7a}$$

This equation shows a similarity, in its dependence on Z, with equation (8). It does not seem possible to decide whether equation (8) would better represent the present observations.

A comparison is made in Table 2 between actual rates of decrease of nuclei concentration as given by the two equations. The values of the two constants in equation (8) corresponding approximately to the conditions of the present experiment are, from NOLAN (1941)

$$\rho = 2 \times 10^{-9} \text{ cm}^3 \text{ sec}^{-1}$$

$$\sigma = 6 \times 10^{-6} \text{ sec}^{-1}$$

The rates of decrease are given for three values of Z. The close agreement for the larger values of Z is probably fortuitous considering the uncertainty in values of the constants but the comparison is, nevertheless, remarkable and supports the interpretation discussed above.

Table 2

Z cm^{-3}	$-dZ/dt$	
	Equation (7*a*) $cm^{-3}\ sec^{-1}$	*Equation* (8) $cm^{-3}\ sec^{-1}$
10^4	$2{\cdot}5 \times 10^{-1}$	$2{\cdot}6 \times 10^{-1}$
10^3	$8{\cdot}0 \times 10^{-3}$	$8{\cdot}0 \times 10^{-3}$
10^2	$2{\cdot}5 \times 10^{-3}$*	$0{\cdot}6 \times 10^{-3}$*

* These values obtained by extrapolation.

VOL. 4 1953

The volume of the gasholder in NOLAN'S experiments was 0·33 m³, whereas the present experiments were performed in rooms 216 m³ and 32 m³ in volume, and under rather different conditions. This indicates that the principal factor, whatever it might be, determining the disappearance of nuclei is a property of the nuclei themselves rather than of the environment.

The importance of the present investigation is not, however, that a method has been found of studying the disappearance of nuclei, but that a new approach has been found for studying the equilibrium of ions in the atmosphere. The presence in the sealed room of an instrument which would measure large-ion concentrations as well as one to measure small ions (or air conductivity) could lead to a complete evaluation of the equation of ion-equilibrium. The complications due to the presence of the apparatus, in this case, on the closed room system could be eliminated by suitable instrumental arrangement. Such experiments may also make possible the quantitative determination of turbulent diffusion coefficients under controlled conditions.

Acknowledgments—The authors wish to thank Professor R. E. HOLZER for many helpful discussions.

The research reported in this paper has been sponsored by the Geophysics Research Directorate of the Air Force Cambridge Research Center under contract AF 19(122)-254.

DL.
4
953

REFERENCES

NOLAN, P. J. 1941 Proc. Roy. Irish Acad. **47,** 25

NOLAN, P. J., and DOHERTY, D. J. 1950 Proc. Roy. Irish Acad. **53,** 163

NOLAN, P. J., and KENNAN, E. L. 1949 Proc. Roy. Irish Acad. **52,** 171

SCHWEIDLER, E. 1918 Wien Ber. **127,** 953

1919 Wien Ber. **128,** 947

Publication No. 35 of the Institute of Geophysics,
University of California at Los Angeles.

Journal of Atmospheric and Terrestrial Physics, 1954, Vol. 4, pp. 322 to 338. Pergamon Press Ltd., London

Radio echoes observed during aurorae and geomagnetic storms using 35 and 74 mc/s waves simultaneously

L. HARANG and B. LANDMARK.

Norwegian Defence Research Establishment, Division of Telecommunication, Kjeller, Norway

(*Received* 11 *November* 1953)

SUMMARY

Observations have been made at Kjeller (near Oslo) and at Tromsoe (near the auroral zone) using pulse transmitters operating simultaneously on 35 and 74 Mc/s. These transmitters were supplied from common pulse modulator and HT units and gave 25 kw output into geometrically similar, rotatable, Yagi aerial systems. The echoes received were displayed on a twin beam oscilloscope with a common time sweep.

The presence of echoes was closely connected with geomagnetic and auroral activity but there was no correlation between the position of the aurora in space and the range of the echoes. Further, when the aerials were pointed at intense auroral forms no echoes were received. Echoes could only be observed when the aerials were directed at a low angle of elevation towards the North. It is concluded that these echoes are not due to direct scatter from the ionosphere or from auroral structures, as assumed previously, but can only be explained as back scatter from land or sea after reflection at an intense E_s layer formed along the auroral zone during aurorae and geomagnetic storms. An analysis of the characteristics of the echoes and the differences in the echo ranges on 35 and 74 Mc/s support this hypothesis. Different types of echoes are described and the results of polarization measurements summarized. Correlation in time between the echo amplitudes is studied by means of double pulse modulation, the time interval between the pulses being variable. Possible passive radiation from the aurora in the 10 cm band could not be detected with the equipment available.

VOL. 4 1953

1. INTRODUCTION

Considerable attention has been paid during recent years to echoes from the ionosphere observed in the VHF-band. In a number of cases echoes have been recorded simultaneously with the appearance of aurorae during geomagnetic storms. Years ago scattering from the ionosphere during a strong auroral display has been observed in Tromsoe (HARANG and STROFFREGEN, 1940) on 41 Mc/s. After the war a number of observers at more southern latitudes have observed echoes in the VHF-band during disturbances. In Manchester, scattered echoes have been recorded on a number of occasions using the echo-recording equipment for meteorite studies. LOVELL, CLEGG and ELLYETT (1947) observing on 46 and 72 Mc/s, recorded echoes partly on 46 Mc/s and partly on both frequencies. The echoes were described as an "increase in the noise level" in the range 450–600 km. ASPINALL and HAWKINS (1950) observing on 72 Mc/s reported scattered reflections during auroral displays. In Canada, CURRIE, FORSYTH, PETRIE and VAWTER (1953) observing in Saskatchewan on 56 and 106·5 Mc/s, recorded persistent and well-defined echoes whenever bright auroral arcs and bands were located in the direction towards which the primary beam was directed. In Kiruna (northern Sweden), HELLGREN and MEOS (1952) have recorded echoes on 30·3 Mc/s using a three-element Yagi antenna rotating in the horizontal plane. A close correlation between the appearance of echoes and simultaneous auroral activity was observed, the extensive material covering one year's observations.

Recently CURRIE, FORSYTH and VAWTER (1953) have given a discussion of their observations from Saskatoon on 56 and 106 Mc/s, covering a period of

observations of almost two years, from 1949 to 1951. The observations were made using antennas rotating in the horizontal plane. The distribution of ranges with azimuth when plotted on an horizontal plane was similar to the distribution given by Hellgren and Meos for Kiruna. CURRIE *et al.* also point out that the ranges of the echoes on the higher frequency 106 Mc/s usually have greater values than the ranges recorded on 56 Mc/s.

2. INSTRUMENTAL EQUIPMENT

In the present investigation all observations of echoes were made on two frequencies simultaneously. The positions of aurorae occurring during the experiments were either observed visually or photographed. Polarization tests were made. In some cases double pulses were transmitted, the time difference between the two pulses being made variable in order to measure the correlation in time between amplitude variations.

The transmitters, one on 35 and the other on 74 Mc/s, had a common HT-supply and common pulse generator. The relative powers and shape of pulses on the two frequencies could therefore easily be maintained constant. The repetition frequency was 50 c/s, locked to the mains. The pulse width could be varied from 8 to 60 μs, usually a pulse width of 23 or 40 μs was used. The pulse generator consisted of two separate pulse channels and double pulses were available. The separation in time between the two pulses could be varied. The antennas consisted of three-element Yagis with a gain 8 db above a dipole (at Kjeller, six-elements), which were mounted on a common frame. The 9 and 4 m antennas thus always had identical field of views. The antennas could be turned in a vertical plane, which was lying in N–S direction. During normal recording conditions the antennas were directed towards north at a low angle. When aurorae appeared the antennas were directed towards the luminous parties of the displays, and it was possible to have in the field of view of the antennas the gradual development of an arc appearing low on the Northern horizon, to bands and draperies with increasing elevations until strong draperies and corona appeared overhead.

During normal recording conditions the plane of polarization of both the transmitting and receiving antennas was vertical. The echoes were displayed on a two-beam oscillograph. Two methods of recording were used. The echoes were either photographed on a fast-moving film allowing each sweep to appear separately, or snaps were taken with one second exposure allowing integration in time of the rapidly changing amplitudes.

The observations were started at Kjeller ($\phi = 59{\cdot}8°$ N, $\lambda = 11{\cdot}1°$ E) in Spring 1952 and during this year echoes were recorded on a number of evenings when great geomagnetic storms appeared. In some cases echoes transmitted at Kjeller were received in Tromsoe both on 35 and 74 Mc/s. In order to make echo observations during visible auroral displays the equipment was transferred to Tromsoe ($\phi = 69{\cdot}9°$ N, $\lambda = 18{\cdot}9°$ E)* and observations were made during the period January-February 1953. During weak geomagnetic storms the aurorae in Tromsoe usually appear low on the Northern horizon at a distance of 300–500 km. During

* The observations were made at the Auroral Observatory, Tromsoe. The authors wish to express their thanks to the Director, Mr. E. TOENSBERG for hospitality and assistance during the observations.

stronger storms the aurorae move closer, may appear in the zenith or even on the Southern part of the sky.

3. Statistics of the Echoes Recorded

The echoes recorded at Tromsoe and Kjeller showed the same main characteristics. The mechanism of reflection at Kjeller and at the arctic station Tromsoe, lying 1120 km to the North of Kjeller, must therefore be regarded as identical. The following features are common to all records: (*i*) the echoes on 35 and 74 Mc/s

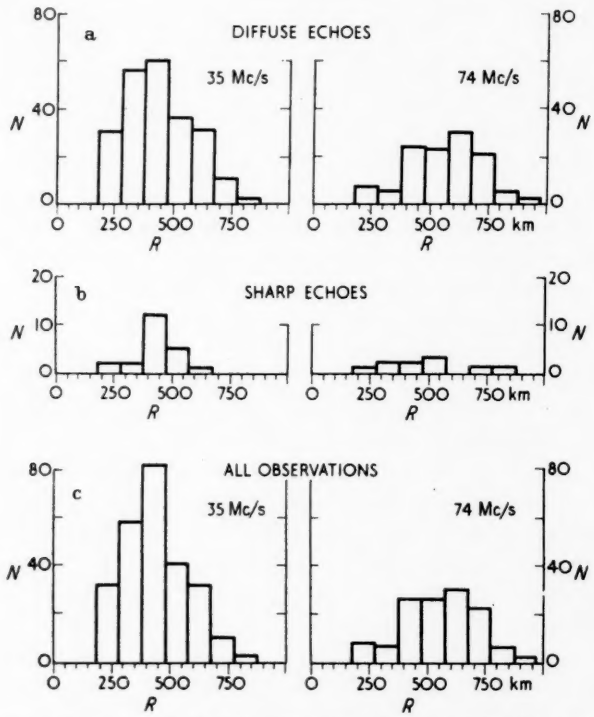


Fig. 2. Ranges of echoes on 35 and 74 Mc/s. The columns in (a), (b) and (c) indicate the numbers, N, observed for each 50 km's increase in range.

have different ranges. The ranges of the 35 Mc/s echoes are always smaller than the simultaneously appearing echoes recorded on 74 Mc/s. (*ii*) there is a distinct difference in character of the echo amplitudes appearing simultaneously on the two frequencies. In Fig. 1 a number of records are given showing the different types appearing.

a. The range of the echoes

In Fig. 2 (a), (b) and (c) histograms of the range, R, as a function of the number of observations, N, are given. The observations have been divided into two groups, diffuse (*a*) and (*b*) sharp echoes. There seems to be no significant difference in ranges between these two types.

VOL. 4 1953

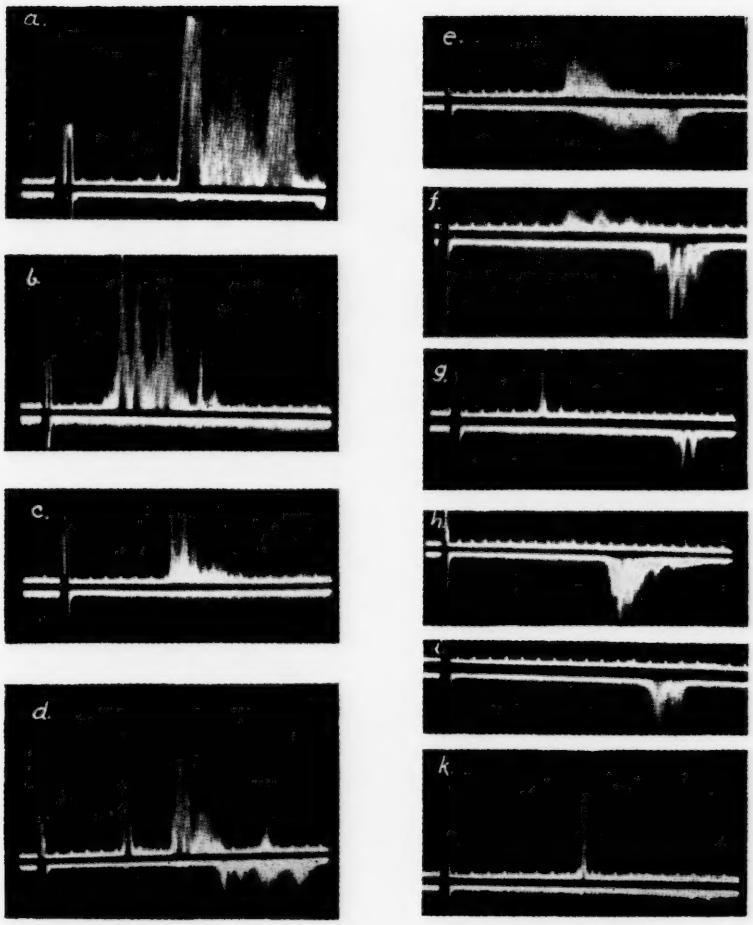


Fig. 1. Echoes on two frequencies received simultaneously on a double-beam oscillograph. The upper trace shows echoes on 35 Mc/s, the lower trace shows echoes on 74 Mc/s. The pips on the upper trace indicate 50 km range marks.

(a), (b) and (c) show echoes on 35 Mc/s only, with leading edge at a range of respectively 220, 240 and 320 km.

(d), (e), (f) and (g) show echoes on 35 and 74 Mc/s simultaneously, the ranges on 74 Mc/s are always greater than the ranges on 35 Mc/s. Record (e) is a typical example of the difference in character of the echoes on 35 and 74 Mc/s. The upper limit of the range of echoes on 74 Mc/s extends to about 1000 km.

(h) and (i) show echoes on 74 Mc/s only.

(k) shows a sharp echo on 35 Mc/s only.

VOL.
4
1953

The histograms show that the most frequent ranges of the 35 Mc/s echoes were in the interval 250–500 km whereas the 74 Mc/s echoes appeared most frequently in the interval 350–700 km. The 35 Mc/s echoes often occurred alone during small magnetic disturbances. In the more intense phase of the storm the 74 Mc/s also appeared but usually with considerably smaller amplitudes. Only in a few cases we have observed 74 Mc/s echoes when 35 Mc/s echoes were absent.

In Fig. 3 the ranges of the echoes are given when echoes on *both* frequencies appear simultaneously. The figure illustrates the characteristic, easily visible on

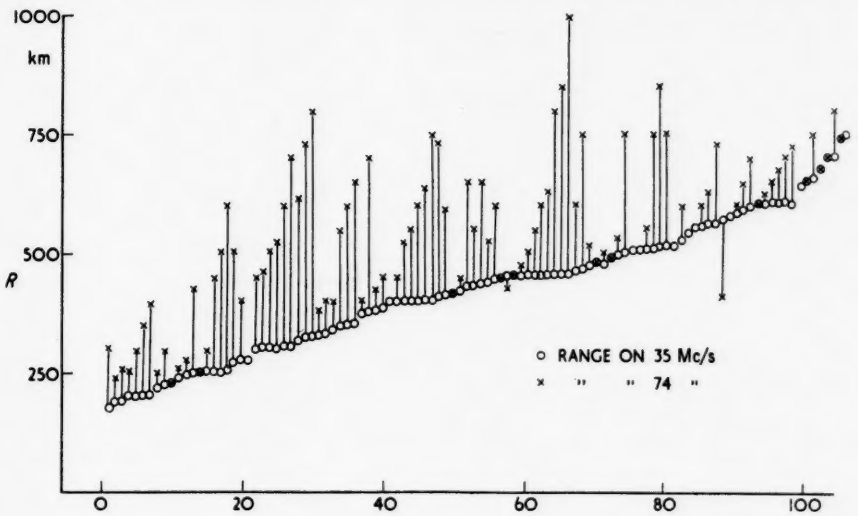


Fig. 3. Ranges, R, of echoes where 35 and 74 Mc/s appear simultaneously. The echoes on 74 Mc/s have a considerably greater range than the echoes on 35 Mc/s.

the records reproduced in Fig. 1, that the 74 Mc/s echoes *always* have a greater range than the simultaneously appearing 35 Mc/s echoes. This difference in range may attain values of several hundreds of km, and in some cases the echoes on the two frequencies appear completely shifted relative to each other on the sweep. This displacement seems to be independent of the mean value of the range. This effect demonstrates conclusively that the scatter observed simultaneously on the two frequencies near the auroral zone cannot be explained as *direct* scatter from the same objects within the ionosphere. An explanation of this difference in range on the two frequencies will be discussed later.

b. The increase in mean range with increasing distance to the auroral zone

HELLGREN and MEOS (1952) and CURRIE *et al.* (1953) have recently published observations which give the magnitude of ranges of the echoes recorded in Kiruna and Saskatoon. It is of interest to compare the mean distribution of the ranges of echoes observed in N–S directions in Tromsoe, Kiruna and Saskatoon with the distances, D_{AZ}, from the place of observation to the auroral zone. In Fig. 4 the distribution of ranges are given for Tromsoe, Kiruna and Saskatoon.

It is evident that there is a gradual increase in range with increasing distance to the auroral zone.

At the top of the figure curves giving the elevation h of the assumed reflection point as a function of range, assuming that the echoes are due to *direct* scatter in an height H equal to respectively 100, 150 and 200 km.

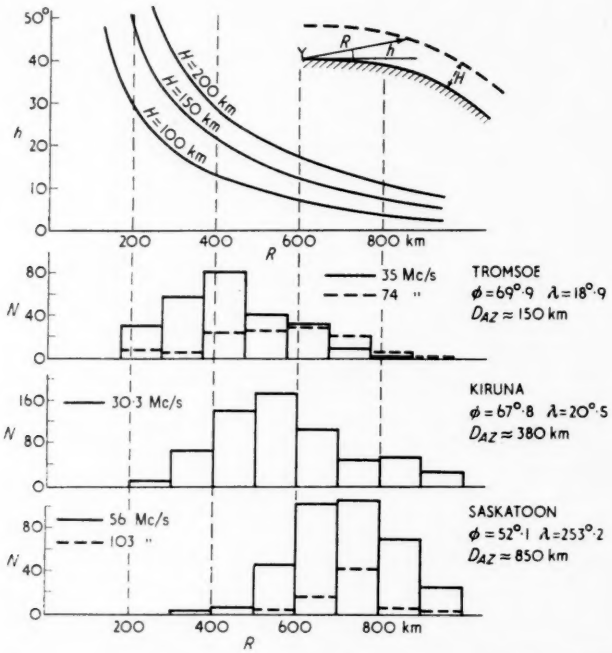


Fig. 4. Increase in ranges of echoes with increasing distance, D_{AZ}, from the place of observation to the auroral zone.

c. The echo duration

ASPINALL and HAWKINS (1950) have distinguished between sharp and diffuse echoes. Our material shows that in far the greater number of cases the scatter appears as diffuse, distorted echoes which may spread over hundreds of km. However, a small number of echoes have been recorded with widths smaller than 25 km. In Fig. 5 the widths of the echoes, ΔR, in km are given in histograms, each column indicating the numbers observed within intervals of 50 km of ΔR. The average widths of the 35 Mc/s echoes are somewhat smaller than those for the echoes on 74 Mc/s. In the first columns of the histograms the sharp echoes have been counted together with diffuse echoes having a smaller width than 50 km. The most frequent value of the width of the echo pattern of 35 Mc/s is 100–150 km, on 74 Mc/s the corresponding value is 150–200 km.

It may be questioned whether this spreading out of the echo pattern actually represents a true variation in range in the direction of the axis of the antenna pattern. The lobes of the Yagis had a half-power width of $\pm 32°$. If now the

echoes entered the lobe uniformly from different directions from a single reflecting or scattering plane, this would introduce a variation in range which could amount to about 15 per cent of the mean range. Now the echo duration usually amounts to a considerably greater value, and we must conclude that the great width of the echo pattern represents a variation in range along the mean direction of wave propagation.

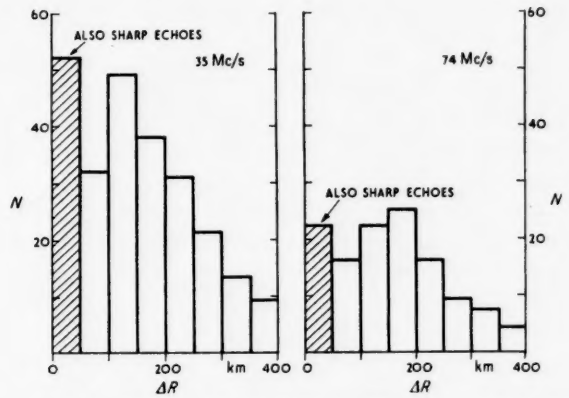


Fig. 5. The widths of echoes, ΔR, on 35 and 74 Mc/s. The columns indicate the numbers, N, observed within intervals of 50 km's increase of width.

d. The amplitude of the echoes

In Fig. 6 the mean values of the amplitudes, A, as a function of range, R, has been plotted. On 35 Mc/s the amplitudes decrease uniformly with range. The decrease in field strength lies between a $1/R$ and $1/R^2$ law, with a closest approach to a $1/R$ law. On 74 Mc/s there is at first an increase in the amplitude with range up to a value of $R = 500$ km, for greater values of R the amplitudes decrease. The explanation of this different behaviour will be discussed later.

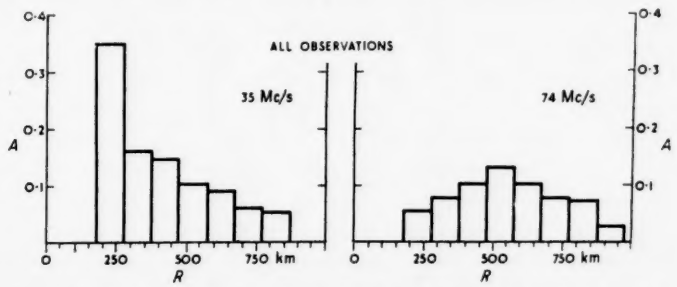


Fig. 6. The amplitudes of the echoes, A, on 35 and 74 Mc/s. The columns indicate the mean value of amplitudes as a function of the range, R.

e. The coefficient of reflection

The VHF-echoes appearing during aurorae and magnetic storms have hitherto been interpreted as due to either reflections from extended, partially reflecting surfaces, or, scattered reflections from "blobs" or stratifications in the ionosphere.

If we assume the simplest model, the waves being reflected normally from an infinite plane surface, the amplitude reflection coefficient ρ is determined by

$$P_2 = P_1 G^2 \rho^2 \left(\frac{3\lambda}{16\pi R}\right)^2 \tag{1}$$

where P_2 is the flux of energy reflected, P_1 is the peak power of the pulse transmitted, G is the gain of the identical receiving and transmitting aerials (in this case folded $\lambda/2$ dipoles with reflector and directors) and R is the range.

Although an alternative explanation of the process of reflection will be put forward in a following section, according to which the echoes are explained as ground clutter reflected back via the E_s-layer, it is convenient to evaluate a coefficient of reflection ρ, using formula (1) from our observations and to compare this with other observations.

A voltage of $V = 100$ μvolt is to be regarded as a representative value of the echo voltage appearing over the input of the receiver. With a peak power of 25 kW and a range of 400 km this gives a coefficient of reflection on 9 m wavelength of $\rho \approx 2 \cdot 10^{-3}$. In some few cases a coefficient of reflection of more than ten times this value has been observed.

The echo amplitudes recorded on 4 m wavelength were considerably smaller than on 9 m. Now when changing the wavelength from 9 to 4 m, the echo amplitudes should decrease in the proportion $(9/4)^2$, if we assume the same value of ρ on both wavelengths. This effect is due to the decrease in collecting area of the aerials. With identical peakpowers and assuming the same coefficient of reflection, we should therefore expect an amplitude ratio of 5 between the echoes on 9 and 4 m. For a large group of echoes this relative decrease in amplitude with wavelength seems to represent the right magnitude.

Now to compare the echo amplitudes on the two wavelengths in order to test quantitatively a mechanism of *direct* reflection from the ionosphere would presuppose that the reflections originated from the same regions within the ionosphere. In Fig. 1 we have demonstrated the difference in range of the echoes on the two wavelengths, an effect which is to be regarded as a general rule. This shift in range is demonstrated through a number of records, which shows that the displacement may range from a small value to a total shift of the echo patterns relative to each other. The difference in character of the echo patterns on the two wavelengths is also evident. On 9 m there is usually a leading edge appearing at the shorter range, on the echo pattern on 4 m this maximum has been displaced towards the distant part of the echo pattern. Further, we have in a number of cases recorded echoes on 4 m without any traces of echoes on 9 m.

If the waves are being reflected or scattered from regions with diffuse boundaries it is well possible that the ranges of the echoes on 4 m will be somewhat greater than on 9 m, corresponding to a deeper penetration of the 4 m waves into the ionized regions before being scattered. It is difficult, however, to conceive a process of direct reflection or scattering from the ionosphere which will explain the total shift of the complete echo pattern amounting to hundreds of km, when changing the wavelength from 9 to 4 m. Further, in some cases we have recorded echoes on 4 m but no echoes on 9 m, an effect which is not easily understood from

VOL. 4 1953

the general formulas for reflection or scattering. These effects force us to give up the picture hitherto accepted of VHF-waves being scattered directly from the ionosphere. In a following section a theory is put forward according to which the echoes recorded are explained as scattering from ground or sea (ground-clutter) reflected back via a strong E_s-layer.

4. Position in Space of Aurorae and Range of Echoes

All observers agree that there is a close *statistical* connection between the appearance of the VHF-echoes and aurorae and geomagnetic storms. Definite opinions have also been expressed (Lovell *et al.* 1947) (Currie *et al.* 1950 and 1953) that the ranges of echoes and direction of antennas indicate that the waves are scattered by individual auroral forms.

Currie *et al.* stress, however, that according to their experience echoes were not obtained from all auroral displays which appeared in the field of view of the aerial structure. Echoes were usually observed with moderately bright aurorae that showed some structure. The echoes rarely persisted for the duration of the auroral display, generally appearing and disappearing intermittently at various bearings during the course of the display. Further, Currie *et al.* report that they never obtained echoes from the Northern sky at altitudes of 16° or higher, and conclude that the echoes have originated exclusively from auroral structures close to the horizon.

Assuming that the echoes were reflected from individual aurorae lying in the centre of the antenna beam, it should be possible to calculate the height of the aurorae if the range is measured and the angular height of the auroral form is measured from a single photo. A curve showing the distribution of heights of the aurorae is given by Currie *et al.*, which fits well in with the well-known height curve of the aurorae. The aurorae measured with this method were appearing at a distance of 700–800 km, and the height above the horizon would therefore be only 5°–10°. According to our experience, however, determinations of height and position of such distant aurorae will not be very accurate.

Hellgren and Meos (1952) do not state if they have made a detailed study of the ranges of the echoes compared with the distances to the auroral forms for individual cases. However, they use the term "localization of aurorae with radar technique," and are apparently of the opinion that the echoes observed are due to direct scatter from auroral forms.

During a number of evenings with great auroral displays in Tromsoe the aerials were directed against various strong auroral forms. The aerials could be turned in a vertical plane in N–S direction. Thus the lobe could be directed towards auroral arcs appearing low in the North. draperies appearing in 30°–60° height on the Northern sky and strong bands and draperies crossing the zenith. In some cases the antenna was also directed against aurorae appearing on the Southern sky. From visual observations and simultaneous photos taken the distances to the auroral forms could be estimated with an accuracy of about 20 per cent.

Compared with Saskatoon the aurorae in Tromsoe appeared over a great range of distances, arcs and bands appearing low in the North at a distance of

500–600 km, to the great aurorae appearing over the zenith in a distance of only 80–100 km. During all these displays echoes only appeared when the aerials were directed towards the North. The ranges of the echoes were usually of the order 300–400 km, with 185 km as the smallest and 1000 km as the greatest range. The beam of the antennas had a half-power width of $\pm$ 32° in the vertical plane. It is therefore not possible to give with accuracy the elevation of the incoming echoes. But the echo amplitudes always showed a maximum when the antenna beam was directed horizontally towards the North, and decreased rapidly with increasing height of the antenna axis. Our observations are therefore in qualitative agreement with CURRIE *et al.* that one does not get echoes reflected from the Northern sky from elevations greater than 16°.

In a great number of cases the antennas at Tromsoe were directed against auroral forms appearing in the sky, from low in the North to the zenith. According to our experience we have not been able to give one example demonstrating a close correlation between the distance to the aurorae visible and the echo range.

One example of this lack of correlation between the two phenomena will be discussed in the following.

On January 29th a geomagnetic storm accompanied by very intense auroral displays occurred during the afternoon and night. The display started with quiet arcs lying in the North at low elevation. The aurorae increased in intensity, moved towards the South, and very intense bands and draperies extending from W to E crossed the zenith. Detailed studies of the correlation between positions of aurorae and appearance of echoes were made. Three examples of the auroral forms recorded are reproduced in Fig. 7 (a), (b) and (c).

VOL. 4 1953

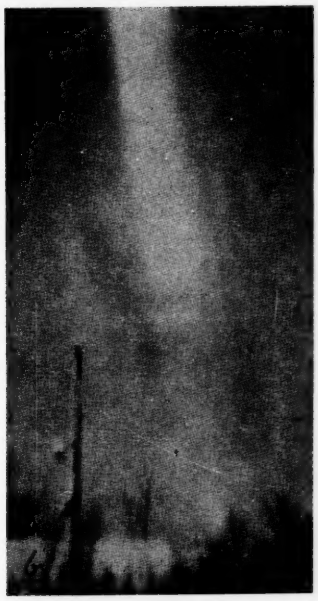
At 17 32 MET a quiet auroral arc (*a*) appeared low in the North at an angular height of 10°–15°, this would correspond to a distance of 350–400 km. Echoes on 35 Mc/s were recorded earlier in the afternoon, but during the appearance of this arc *no* echoes could be detected when the antenna was directed towards the central part of the arc.

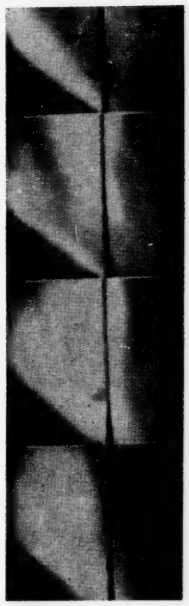
At 18 27 MET very intense bands and draperies (*b*) appeared crossing the zenith from W to E. The antenna was directed towards the very luminous parts of the zenith. No traces of echoes could be observed although the distance to the lower edge in this case amounted to only 80 km and the maximum luminosity was at a distance of 110–120 km.

At 19 10 hrs a similar very intense drapery (*c*) crossed the zenith from W to E. In this case, also, no echoes appeared when the antenna beam was directed towards the very luminous parts at the zenith. However, *strong echoes* now appeared on 35 Mc/s when the antenna was pointed low towards the horizon in the North where no aurorae were visible.

Previous explanations of the scattering of VHF-waves during auroral displays have assumed that the scattering process was similar to the scattering from meteorite trails (HERLOFSON, 1947). Scattering from meteorites exhibits a strong maximum of amplitude when the waves impinge normally on the trail of a meteorite. Now the direction of the streamers in the aurorae lie along the earth's magnetic field. The echoes from the aurorae should therefore be strongest when the antenna is directed towards luminous parts of the aurorae having streamers

(a)





(b) (c)

Fig. 7.

(c) Auroral are appearing low in North at a distance of 350-400 km. No echoes were recorded.
(b) Western part of a very intense band-drapery, crossing the zenith. Distance to the luminosity in zenith 80-110 km. No echoes from the direction of the zenith were recorded.
(c) Film strip of western part of very intense drapery, crossing the zenith. No echoes from the direction of the zenith, but echoes appeared from the Northern horizon where no aurorae were visible.

VOL.
4
1953

perpendicular to the direction of wave propagation. Using this assumption, CHAPMAN (1953) has computed curves giving the ranges of echoes to be expected for waves impinging normally on auroral streamers lying at various heights above the earth's surface when viewed from different observing places.

In the observations here described the waves were in no cases impinging at right angle to the direction of the earth's magnetic field. For the example shown in Fig. 8 the waves would in the case of the diffuse arc (*a*) appearing on the Northern sky, impinge at an angle of 12° from the desired ideal direction. In the two cases (*b* and *c*) where the brilliant draperies with streamers appeared in the zenith, the direction of observation was almost along the earth's magnetic field.

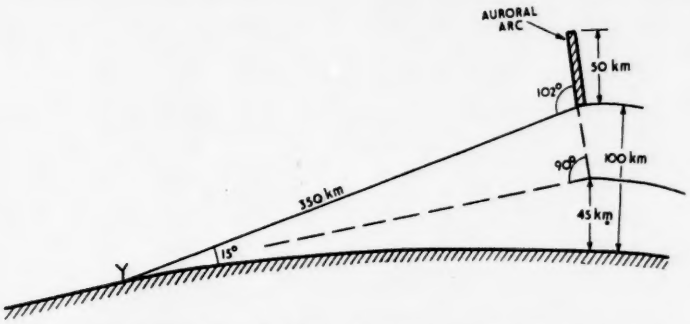


Fig. 8. Position of arc relative to the directions of wave propagation.

The observations discussed in the examples *a*, *b* and *c* have been confirmed in a number of cases. Diffuse auroral arcs lying low on the Northern sky give no reflections, although the waves may impinge almost at right angle to the main direction of the field, the deviation may be only 15°–25°. Draperies and bands with ray-structure appearing overhead near the magnetic zenith, where the waves will impinge along the direction of the field, also give no reflections, even during the most intense displays in a distance of only 80–100 km.

5. Appearance of VHF-echoes Simultaneously with Increase in Critical Frequency of the E_s-layer during a Geomagnetic Storm

During the observations in Tromsoe ionospheric soundings in the range 0·5–25 Mc/s were usually taken each half an hour, thus giving the critical frequencies of the ionospheric layers. In a number of cases it was observed that strong scatter on 35 and 74 Mc/s appeared when the critical frequency of the E_s-layer attained high values. In Fig. 9 the appearance of VHF-scatter, critical frequencies of the E_s-layer and the variations of the earth's magnetic field, H-component, are given for three evenings. It is evident that the VHF-echoes appeared only during the period with increased values of E_s-ionization.

Analysis of the observations shows that scatter on 35 Mc/s may appear even during small storms, whereas the echoes on 74 Mc/s are restricted to the more severe phase of the storms. The intimate connection between high critical frequency of E_s and appearance of VHF-echoes indicates that the echoes are to be

explained as scatter from sea (or land) reflected back via the sharp lower boundary of the E_s. When the waves are transmitted horizontally an MUF-factor up to 5 is to be expected. The maximum values of critical frequencies of the E_s on the evenings mentioned in Fig. 9 should therefore allow reflection of waves up to

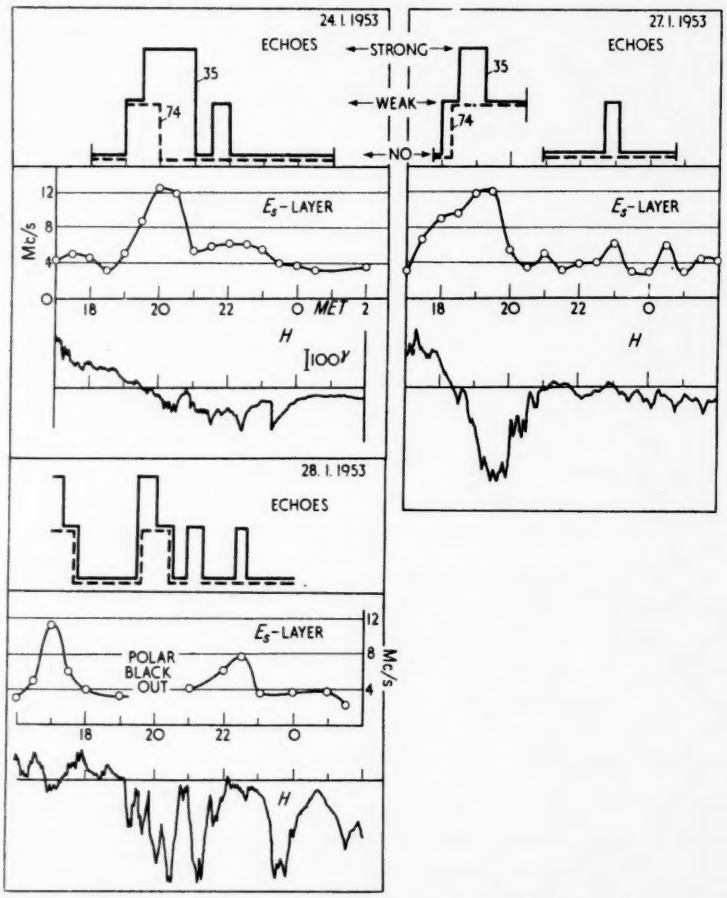


Fig. 9. Appearance of echoes on 35 and 74 Mc/s and periods of high critical frequencies of the E_s layer during a geomagnetic storm, recorded at Tromsoe. The horizontal intensity, H, of the earth's magnetic field is shown below.

60–65 Mc/s. Now the E_s is recorded vertically in Tromsoe, whereas the assumed reflection point will be lying at a distance 150–350 km to the North, closely under the normal auroral zone, where one may assume E_s ionization to be still higher.

6. Process of Reflection

As previously stated the effect of different ranges of the echoes on 35 and 74 Mc/s force us to give up the picture hitherto accepted of VHF-waves being scattered directly from the ionosphere. Further, the appearance of the echoes are not correlated with the appearance of visual auroral forms, but with the appearance

VOL. 4 1953

of a strong E_s-layer. Instead of being directly reflected from the ionosphere the VHF-echoes are explained as back scatter from the ground or sea reflected back via the strong E_s-layer. In the following we will discuss the character of the echoes from the point of view of being back scatter. Further, two models of the reflection process will be discussed which may explain the shift in ranges from 35 to 74 Mc/s.

a. The character of the echoes

Backscatter of radiowaves in the usual HF-band has been extensively studied since ECKERSLEY's (1940) first discussion of the effect in 1932. In the HF-band the most frequent scatter process is backwards reflection of the waves over the skip zone via the F_2-layer. A common feature of this scatter is a pronounced focus effect. The scatter usually exhibits a sharp leading edge followed by a maximum value of the amplitude and slowly decreasing amplitudes towards greater range. The focus effect is due to the addition of high and low angle rays in a layer with considerable, vertical extension. If, however, the waves are reflected back via thin, homogeneous layers where the reflections are partial, no focus effect will appear. The echo amplitude should in this case increase uniformly with range due to the increase in MUF-factor. From the shape of the echoes it should thus be possible to decide whether the reflections come from an extended layer, or from partial reflections in thin irregular horizontal stratifications in the layer.

The echoes recorded were accordingly divided into four groups according to the position of the maximum amplitude within the scatter. In group I the maximum amplitude is near to the leading edge and within the first quarter of the total echo duration. In the groups II, III and IV the maximum amplitudes occurred in successive quarters of the echo duration. The number of echoes appearing in each of these four groups were counted. The 35 Mc/s were further divided into weak, medium and strong echoes. The 74 Mc/s echoes were grouped with ranges greater or smaller than 600 km. The results are given in Table 1.

Table 1. *Position of maximum amplitude within the echo form*

a. 35 *Mc/s*

Group	*Weak*	*Medium*	*Strong*	%
I	46	33	37	57
II	34	15	10	29
III	21	4	2	13
IV	2	1	0	1
Total	103	53	49	205

b. 74 *Mc/s*

Group	*Echo range* $\leqq$ 600 km	*Echo range* $>$ 600 km	%
I	14	10	26
II	8	7	17
III	13	8	23
IV	27	4	34
Total	62	29	91

Table 1 *a* shows that on 35 Mc/s the maximum amplitude in 86 per cent of the cases appears in the first half of the echo duration. These echoes demonstrate the focus effect in the back scatter, and the reflection process must be regarded as

coherent refraction in an extended E_s-layer. 14 per cent of the echoes may, however, be regarded to have been partially reflected in a thin layer. These echoes were usually weak.

Table 1 *b* demonstrates the different character of the 74 Mc/s echoes. The main tendency is here to have the maximum amplitude in the distant part of the echo form. This effect is especially pronounced for the 74 Mc/s echoes with ranges less than 600 km and may therefore be explained as due to the MUF-factor. For great ranges the MUF-factor is nearly constant and the effect should be absent.

The conclusion is that the reflection of the 74 Mc/s waves must, in general, be a partial reflection via a thin layer. The anomalous variation of amplitude with range, which is evident in Fig. 2, may be explained as due to the MUF-factor.

The general conclusion is therefore that the scatter on 35 Mc/s exhibits all the characteristics which have been described previously for backwards scatter by ECKERSLEY. In some cases, even two-hop reflections of the 35 Mc/s scatter via refraction in an extended E_s-layer seem to occur (see Fig. 14).

For the 74 Mc/s echoes the reflection process must be different, and the principal fact to be explained is the systematic difference in range between the 35 and 74 Mc/s scatter. We assume that this effect must be explained through the reflection mechanism in the E_s-layer. As highly idealized models which will give a shift in range of reflections on 9 and 4 m wavelengths, the following two conditions of the E_s-layer will be discussed.

VOL. 4 1953

b. Partial reflection at a thin layer within an extended E_s-layer

The extended E_s-layer has sufficient electron density for regular refraction of 9 m waves at an angle giving a skip zone corresponding to the range of the 9 m echoes observed. At this angle of incidence, however, the E_s-layer is not sufficiently dense for refraction of the 4 m waves. At a lower angle of elevation, α_4, the 4 m waves may be partially reflected from a thin layer within the E_s-layer, and the echoes on 4 m will have a greater range than the refracted waves on 9 m, see Fig. 10 (a).

This model of E_s leads to the conclusion that, even during the strongest auroral displays and geomagnetic storms the extended E_s is not sufficiently dense to allow refraction of 74 Mc/s even at the lowest angle of elevation where the MUF-factor has a value of about 5. This gives an upper limit of the critical frequency of E_s attainable of about 15 Mc/s.

c. Partial reflection in circular "ion-plates" with dimensions corresponding to one Fresnel-zone

We will illustrate the effect through a numerical example. At vertical incidence the radius of the first Fresnel zone $a = \sqrt{\lambda R}$ for $\lambda = 9$ m and $R = 100$ km will be 0·94 km. For a distance $D_9 = 500$ km, which will give an angle of elevation of $\alpha_9 = 22{\cdot}4°$ the zone plate will have a radius of 4·1 km. For $\lambda = 4$ m this plate, however, will act as a zone plate at a lower angle of elevation, $\alpha = 17{\cdot}3°$ and we get a distance $D_4 = 700$ km. It is thus evident that, using this model of reflection process, we may get a total shift of the echo patterns on 35 and 74 Mc/s relative to each other, (see Fig. 10b).

In model (*b*) we have assumed a thin layer within the E_s-layer, in model (*c*) we have further assumed that this stratification is divided into circular "blobs." Which of these models corresponds to the actual conditions in the ionsphere is difficult to decide, but we are inclined to assume that the case (*b*) is the more probable one.

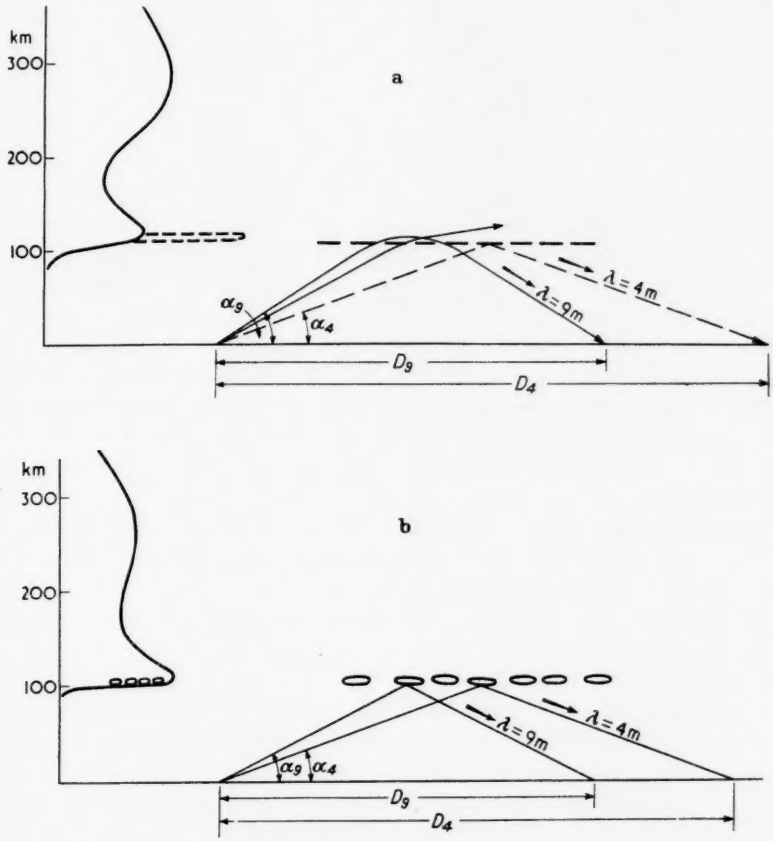


Fig. 10 (a) and (b).

7. Polarization Tests

During all observations the elements of the transmitting aerial systems were vertical. The position of the receiving aerials were either vertical or horizontal. In some tests we used two separate receiving channels, the one with vertical and the other with horizontal Yagis. Both echo patterns were then displayed for comparison on the two traces of the double beam oscillograph.

The results of the tests on 35 Mc/s were as follows:

(*a*) echoes with great amplitudes and with a leading edge: no change in amplitude from vertical to horizontal receiving antennas, see Fig. 11,

(*b*) sharp and narrow echoes: a polarization effect was observed. The amplitude with receiving antenna vertical (parallel to the transmitting antenna) was stronger than with receiving antenna horizontal,

(*c*) more diffuse echoes without leading edge: same results as in (*a*).

On 74 Mc/s only a few tests were made. In some tests the echoes on 74 Mc/s only appeared when the transmitting and receiving antennas were parallel.

The 35 Mc/s echo patterns with great amplitude and a leading edge, which include by far the greater number of echoes observed, show no polarization effects. The sharp echoes on 35 Mc/s and the echoes on 74 Mc/s have a tendency to preserve their plane of polarization.

8. FADING OF THE ECHOES

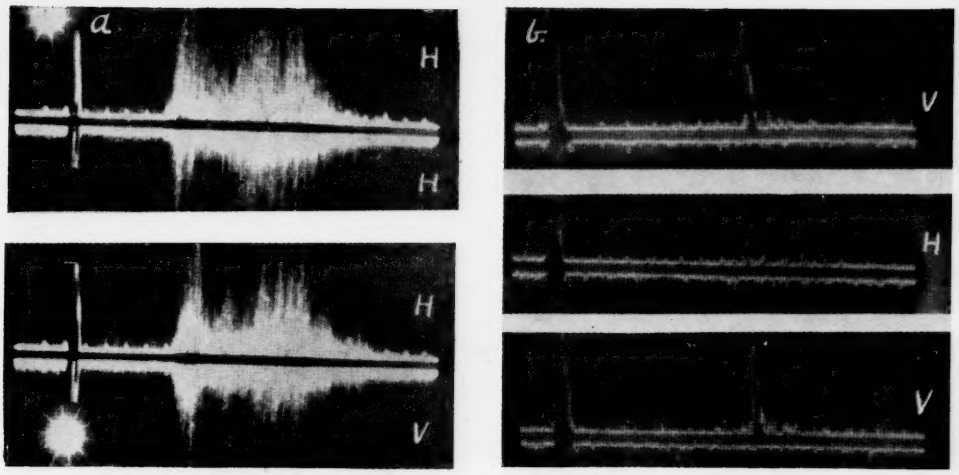
Although the scattered echo may preserve its mean amplitude and character for some time, there is always a very rapid fading within the echo—the echo is "boiling." Records on a rapidly moving film show that the echo pattern has changed completely from one sweep to the next. There is thus no correlation in amplitude between sweeps taken 1/50 sec. apart. In order to investigate this point more closely double pulses were transmitted. The time difference between the two pulses could be made variable. A number of tests were made, and it could be shown that the amplitudes maintained their character up to a time difference of 3–4 $\times$ 10^{-3} sec which should represent the magnitude of the autocorrelation function. With greater time differences between the double pulses the echoes changed character.

This should correspond to a Doppler change of about 250 c/s. MORE (1951) reports that VHF-signals reflected back during auroral displays are characterized by an extremely high rate of fading, which renders voice modulation unintelligible. This is in accordance with the high fading rate evident from our echo records.

9. MOVEMENT OF THE E_s-LAYER

Records and visual observations of the scatter show a characteristic variation of the range with time. When the scatter first appears with small amplitude the range has usually its maximum value. With increasing amplitude the range decreases and usually has its maximum value at a minimum value of range. When the amplitude again decreases the range increases until the echo fade away often rapidly. There is thus a continuous change in the position of the part of the E_s-layer which acts a reflecting screen. A change in the range of the echoes of about 100 km/min is a value which often may appear. This movement of the reflecting screen is not to be regarded as a bodily drift of a uniformly ionized cloud. We have to assume that new ionized parts of the cloud is continuously being formed due to intrusion of ionizing particles. The resulting picture will be that of a stationary ionized cloud, the outer limit of which is continuously expanding or contracting in various directions. Such a pulsating cloud will introduce rapid fading of the main signal, due to interference between the reflected signals from the stationary part of the cloud and the pulsating edges. If we assume that the edge changes its position with a velocity of 2 km/sec this will introduce a phase-change of about 220 λ between the pulsating and quiet part of the cloud

VOL. 4 1953



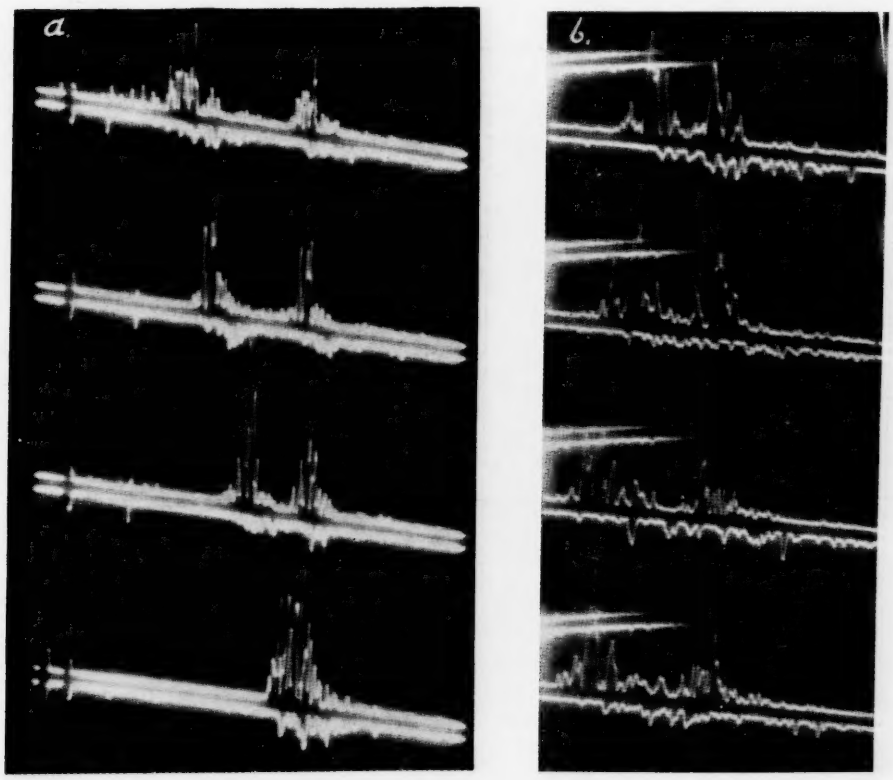
(a) (b)

Fig. 11. Polarization tests. Transmitted wave vertical.

(a) Echoes on 35 Mc/s, great amplitudes and leading edge, received through *two* separate channels, with directions of receiving antennas vertical-vertical and vertical-horizontal. No polarization, effect.

(b) Echoes on 35 Mc/s, sharp echoes. Single sweeps of echo patterns with receiving antenna vertical, horizontal and vertical. There is a tendency to preserve the plane of polarization.

DL.
4
953



(a) (b)

Fig. 12. Sweeps of echoes on a rapidly moving film, time difference between each sweep 1/50 sec. there is no correlation between the amplitudes from one sweep to the next. Double pulse modulation with changing time difference between the pulses is used. In (b) the time-base is expanded. The correlation between the amplitudes diminishes strongly when the time interval between the pulses exceeds $3—4 \times 10^{-3}$ sec.

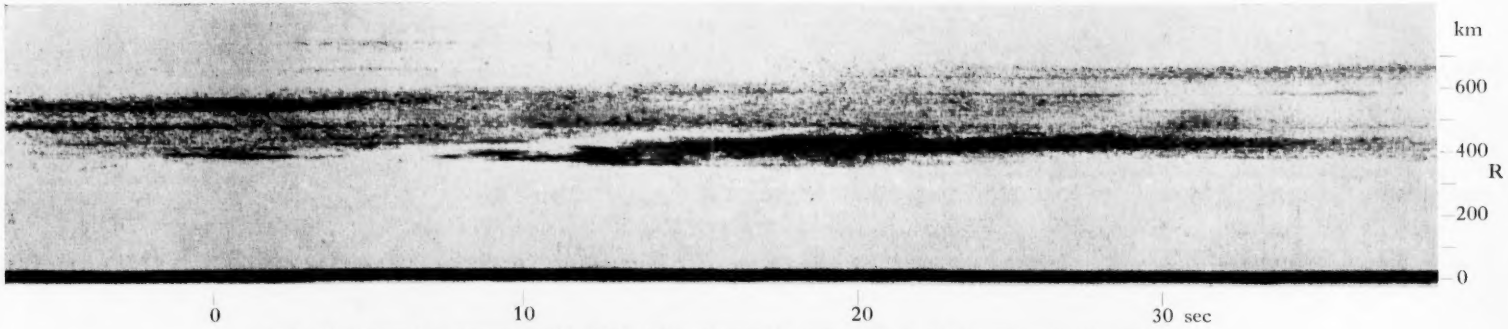
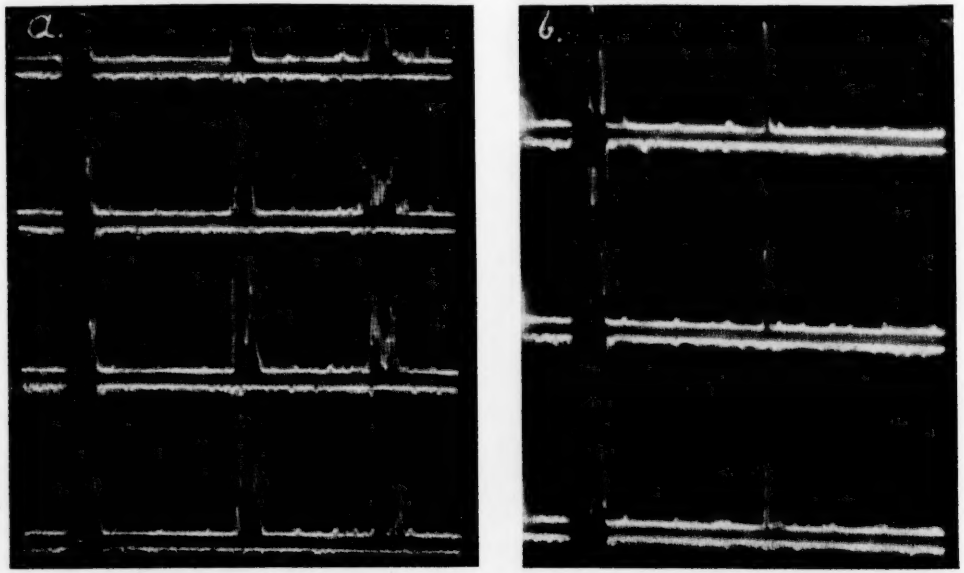


Fig. 13. Record of scatter on 35 Mc/s. The change in range, R, is of the order 2—10 km/sec.



(a) (b)

Fig. 14.

(a) Double hop scatter on 35 Mc/s.
(b) Sweeps of reflection from a meteorite.
The scatter shows a complete change of echo amplitudes from one sweep to the next, the meteorite echo shows a continuous change in amplitude.

VOL. 4 1953

on a wavelength of $\lambda = 9$ m. We thus get a fading frequency of about 220 c/s. This value fits in with the magnitude of the autocorrelation function of about 4×10^{-3} discussed in section 8.

10. Multiple Echoes

In a number of cases the records of the 35 Mc/s scatter show traces of multiple scatter, a second echo group appears having a range about twice the value of the first strong echo group. Such multiple back scatter via the F-region has previously been observed on lower frequencies. Villard and Peterson (1952) have demonstrated this double hop back scatter on 14 Mc/s. Fig. 14 shows a record demonstrating double hop back scatter on 35 Mc/s. The range was in this case 185 km for the first hop. No echoes on 74 Mc/s appeared. In this case the E_s-layer must have had a great extension in N–S direction. No double hop scatter on 74 Mc/s has been observed.

11. Pulse Transmissions, Kjeller–Tromsoe

In the summer and autumn 1952, double pulses were transmitted at Kjeller and received at Tromsoe on a number of evenings. The base line was 1120 km. The pulses appeared in Tromsoe only on evenings when great geomagnetic storms occurred, and then only during the most perturbed phase of the storm. The pulses on 35 Mc/s were frequently recorded, on 74 Mc/s only during one evening. The pulses received at Tromsoe were undistorted having the same width of 40 μsec as the transmitted pulse, whereas the back scatter received at Kjeller always showed the usual severe distortion. Also the rate of fading of the pulses received at Tromsoe, estimated visually, was slow and resembled normal echo fading. This reception of undistorted pulses is to be expected, since both a forwards scatter process or the usual ionospheric refraction process will give negligible distortion.

12. Passive Radiation from Aurorae

Forsyth, Petrie and Currie (1950) have recently announced that, during auroral displays, radiation in the 3000 Mc/s band has been received when the antenna-paraboloid was directed towards the luminous surfaces of the aurorae. The radiation occurred as narrow pulses or bursts arriving in random manner and lasting only a small fraction of a second. Each pulse was estimated to be of one to five μ sec duration. Observations on twenty auroral displays seem to confirm the auroral origin of these pulses.

At Tromsoe tests were made in order to repeat these observations. A 10 cm link receiver* was used. The receiver sensivity was 14 db below $kT\Delta f$, $\Delta f = 4{\cdot}5$ Mc/s. The antenna was a paraboloid of diameter 6 feet, the gain was 32 db above an isotropic radiator. The noise was displayed on an A-scan of a CRT with 5 Mc/s bandwidth.

* The receiver and aerial were kindly put at our disposal by the Radar Division of the Defence Research Establishment. The authors wish to express their thanks to the chief scientist, Mr. Sture Koch and to Messrs. Sörbye and Olesen for assistance during control measurements of the receiver arrangement.

During strong auroral displays the antenna was directed towards the luminous parts of the aurorae, careful tests were made against a series of very brilliant bands and draperies appearing in the zenith at a minimum distance of 80–100 km. No pulses or bursts, however, could be observed which exceeded the normal noise level on the CRT-trace.

We must point out that the observing conditions in Canada and Tromsoe may not have been quite identical. In Tromsoe the observations on the *strong* aurorae were made in a direction almost along the earth's magnetic field, or only slightly inclined to this direction. In Canada the observations were apparently made at about right angles to the direction of the earth's magnetic field. Observations in Tromsoe on intense but *distant* aurorae at the Northern horizon, lying at a distance of 350–400 km did not, however, show any radiation, although in this case the direction of observation was almost at right angles to the field of force.

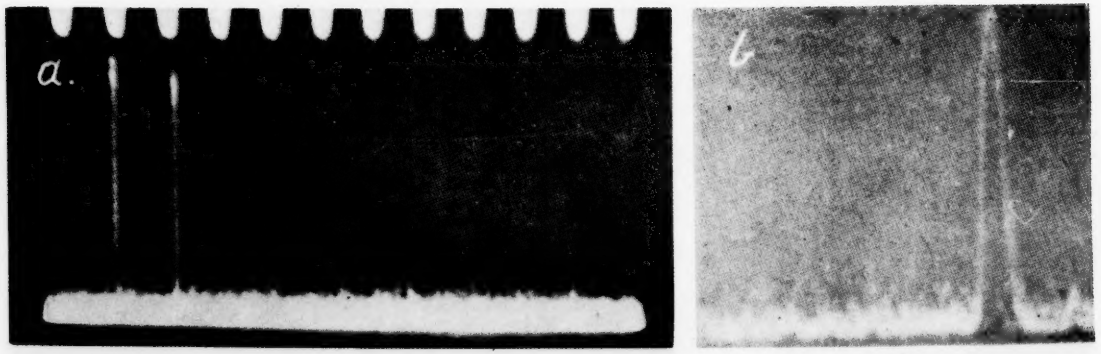
Acknowledgments.—The authors wish to express their sincere thanks to Mr. A. CROWO for his most valuable assistance during the development of instruments and carrying through the observations. During the observations in Tromsoe we have been most ably assisted by Mr. REIDULF LARSEN.

We also wish to express our sincere thanks to Mr. F. LIED and Dr. N. HERLOFSON for valuable discussions during the analysis of the observations.

REFERENCES

ASPINALL, A. and HAWKINS, G. S.	1950	J. Brit. Ast. Ass. **60**, 130
CHAPMAN, S.	1953	J. Atmosph. Terr. Phys. **3**, 1
ECKERSLEY, T. L.	1940	J.I.E.E. **86**, 548
CURRIE, B. W., FORSYTH, P. A. and VAWTER, F. E.	1953	Nature **171**, 352 J. Geophys, Res. **58**, 178
FORSYTH, P. A., PETRIE, W. and CURRIE, B. W.	1949 1950	Nature **164**, 533 Canad. Journ. of Res. *A*, **28**, 424
FORSYTH, P. A., PETRIE, W., VAWTER, F. and CURRIE, B. W.	1950	Nature **165**, 561
HARANG, L. and STOFFREGEN, W.	1940	Zs. f. Hochfrequenztechn. **55**, 105
HELLGREN, G. and MEOS, J.	1952	Tellus **4**, 249
HERLOFSON, N.	1947	Nature **160**, 867
LOVELL, A. C. B., CLEGG, J. A. and ELLYETT, C. D.	1947	Nature **160** 372
MORE, R. K.	1951	J. Geophys. Res. **56**, 97
VILLARD, V. G. and PETERSON, A. M.	1952	QST, **56**, 11 (March)

GEOPHYS. INST. BERGEN
VOL. 4
1953



(a) (b)

Fig. 15. Pulses on 35 Mc/s transmitted from Kjeller and received in Tromsoe during strong geomagnetic storm. The pulses received in Tromsoe are undistorted. (a) shows the double pulse, (b) the time-base is triggered by the first pulse in order to show details of the pulse on an extended time-base.

VOL.
4
1953

RESEARCH NOTES

The effect of the horizontal extent of Aurorae on the apparent luminosity-altitude curves

(*Received* 13 *November* 1953)

§ 1. The identification of lines of atomic hydrogen in auroral spectra by VEGARD (1939) and their subsequent study by GARTLEIN (1950) and MEINEL (1950) show that fast protons must certainly be a component of the incoming stream of primary particles. Furthermore the recent quantal calculations of BATES and DALGARNO (1953) indicate that the electron capture process presumed to be operative is, in fact, a very inefficient source of the lines, which suggests that fast protons may actually be a major component.

It is a simple task to derive the luminosity-altitude distribution that would ensue from the entry of protons into the Earth's atmosphere. Except perhaps in the case of the higher aurorae the calculated curves for a mono-energetic incident stream (cf. BATES and GRIFFING 1953) are very much sharper than are the curves obtained from photometric measurements (cf. HARANG 1944, MEINEL 1952, BERKEY and GARTLEIN 1952); in particular it has been shown that the hydrogen line emission from an aurora located near the 100 km level would be confined to a vertical interval of only 1 or 2 km, which is a far narrower range than the observations indicate. BATES and GRIFFING (1953) have pointed out that the predicted curves could be reconciled with the measured curves by arbitrarily assuming that the incoming protons have a wide energy spread. However there is the possibility, to which MEINEL (1952) has drawn attention, that the finite horizontal extent of aurorae along the direction of the magnetic meridian may conceal the true luminosity-altitude distribution. It is the purpose of the present note to investigate this effect.

§ 2. For convenience we shall consider a two-dimensional model aurora throughout which the emission per unit volume is simply

$$L(z, x) = \mathscr{L} z \exp\left(\frac{-z}{\alpha} - \frac{4x^2}{\beta^2}\right) \quad (z, +\text{ ve})$$

$$= 0 \qquad\qquad (z, -\text{ ve}) \tag{1}$$

where z is measured vertically from the base which is at an altitude h_0, x is measured horizontally from a line making an angle θ with the Earth's surface, α and β are parameters defining the vertical and horizontal spread of the aurora and $\mathscr{L}$ is a proportionality constant (see Fig. 1). Clearly the true luminosity-altitude distribution is given by

$$L(z) = \int_{-\infty}^{\infty} L(z, x)\, \mathrm{d}x$$

$$= \mathscr{L} \frac{\sqrt{\pi}}{2} \beta z \exp\left(\frac{-z}{\alpha}\right). \tag{2}$$

To an observer at a point O it would appear that the luminosity in the direction of a vector $\mathbf{s}$ making an angle ϕ with the horizontal is

$$L(\phi) = \int L(z, x)\, ds$$

$$= \mathscr{L} \int_0^{\infty} z \exp\left(\frac{-z}{\alpha} - \frac{4x^2}{\beta^2}\right) ds. \tag{3}$$

This expression may readily be evaluated with the aid of any of the standard tables of the probability integral. In order to convert the luminosity-angle curves thus obtained into luminosity-altitude curves it is of course necessary to associate with each angle a definite altitude. Unfortunately the association is rather ill-defined—a fact that has not been taken properly into account in the presentation of the observational data in the literature. It is, however, convenient and natural to take the altitude h corresponding to a given angle

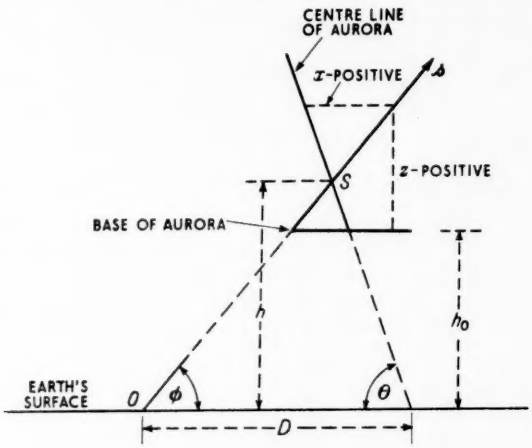


Fig. 1.

ϕ to be that of S, the point of intersection of the observing line and the central line of the aurora. Adopting this convention calculations were carried out for the following cases. (see Fig. 1.)

Angle of central line	θ	75°
Altitude of base	h_0	100 km
Vertical extent parameter	α	1 to 20 km
Horizontal extent parameter	β	0·5 to 200 km
Observation distance	D	200 km and 600 km

§ 3. The luminosity-altitude curves may be characterized (HARANG 1944) by the distances $t_l^n(\alpha, \beta)$ and $t_u^n(\alpha, \beta)$ between the level of maximum intensity and the two levels at which the intensity is a fraction, n, of the maximum, the subscript l referring to the level in the lower part of the aurora and the subscript u to the level in the upper part. For aurorae with bases near the 100 km level HARANG (1944) reports that $t_l^{0.5}$ and $t_u^{0.5}$ average about 6·8 km and 13·5 km respectively. The observed luminosity curves of such aurorae may be represented by equation (2) if α is taken to be about 8 km.

From the computations which were carried out it was found that the observed and true values of $t_l^n(\alpha, \beta)$ and $t_u^n(\alpha, \beta)$ are almost the same except when β/α is appreciably greater than unity. Consequently the apparent luminosity-altitude distribution of HARANG, with α as indicated above, may be accepted as the actual luminosity-altitude curve unless β is some 10 km or greater (which seems unlikely to be common).

It was found also that if the emission were confined to a vertical interval of 1 or 2 km (cf. § 1) it would be necessary for β to be extremely large to yield apparent luminosity-altitude curves such as are observed, the required value, which of course depends on the observing distance, being some 50 km when D is 200 km and some 100–200 km when D is

VOL. 4 1953

600 km; moreover the apparent luminosity-altitude curve would then be controlled almost entirely by the value of β and would presumably be very variable. In view of these results it may be concluded that the finite horizontal extent of aurorae cannot be invoked to explain the discrepancy between the measurements and the predictions based on the assumption that the incoming protons are mono-energetic.

Acknowledgements—We wish to thank Professor D. R. BATES for suggesting this problem and for many helpful discussions.

Department of Applied Mathematics,
The Queen's University of Belfast

G. W. GRIFFING*
A. L. STEWART

References

BATES, D. R. and DALGARNO, A.	1953	Proc. Phys. Soc. A. **66**, 972
BATES, D. R. and GRIFFING, G.	1952	J. Atmosph. Terr. Phys. **3**, 212
BERKEY, D. K. and GARTLEIN, C. W.	1952	Mem. Soc. Roy. Sci., Lèige **12**, 199
GARTLEIN, C. W.	1950	Trans. Amer. Geophys. Union **31**, 7
HARANG, L.	1944	Geof. Publ. **16**, No. 6
MEINEL, A. B.	1950	Astrophys. J. **111**, 555
MEINEL, A. B.	1952	Mem. Soc. Roy. Sci. Liège **12**, 203
VEGARD, L.	1939	Nature **114**, 1089

A note on solar-terrestrial relationships

This note suggests a possible explanation for some of the deviations obtained when correlating particular solar events with geophysical disturbances presumably caused by corpuscles ejected from the sun.

It is generally agreed that aurorae, geomagnetic variations, etc., are caused by the impingement of solar particles upon the earth's atmosphere. The particle stream is essentially neutral and consists of approximately equal numbers of electrons and positive ions. Neutral particles also exist in equilibrium with the radiation field, but their number is relatively small. The approximate charge equality may occur on emission from the sun, by acquisition of the deficient charges from the ions of interplanetary space, or by other means.

Several theories have evolved to explain the acceleration of the solar particles towards the earth's dark hemisphere. STÖRMER mathematically described the trajectory of a single charge as it approaches a magnetic dipole. Depending upon their velocity vectors, charges approaching the earth could strike the sunlit hemisphere, the dark side, or could be entirely deflected from the earth. CHAPMAN and FERRARO objected to the application of STÖRMER's results to auroral theory, mainly on the basis that the treatment of only isolated charges moving in a magnetic dipole field did not represent the probable conditions involved in auroral formation and geomagnetic changes. Their model envisages a neutral stream (or plasma) of charged particles which after leaving the sun engulfs the earth. Some aspects of the geomagnetic changes appear to be clarified by their model. However, it is not quite clear how particles in the "hollow" behind the earth leave the stream to invest the dark hemisphere. No consideration is given to the possibility that charged particles may strike and cause aurorae in the sunlit hemisphere. Of the two theories, the former is considered to apply to cosmic rays and the latter to aurorae and magnetic disturbances.

* Affiliated to the Geophysics Research Directorate of the U.S. Air Force, Cambridge, Massachusetts

This note proposes that corpuscular emission from the invisible solar disk may be deflected by the sun's general magnetic field and directed towards the earth. Once approaching the earth, it may strike the dark or sunlit side depending upon energy considerations. These particles may be responsible for the many instances of aurorae and magnetic disturbances which do not correlate with activity observed on the visible solar 341 disk. This mechanism would weaken daily correlations between observed solar activity on the visible disk and geophysical disturbances, but at the same time could permit monthly correlations to be strong. For example, the correspondence between changes on the earth (or in its atmosphere) and an observable solar event could be very low, or zero, if the solar causative agent occurred behind the sun. On a monthly basis, however, as solar activity increased so would the geophysical effects. Present studies of solar-terrestrial effects are biased towards the observable sun.

The mechanism responsible for deflecting ejecta from behind the sun to the earth is the same as that forcing charged particles to the earth's dark hemisphere. All charged particles leaving the sun would interact with the general and local magnetic fields. However, the solar magnetic dipole is so enormous when compared with the local fields, that its effects are by far the stronger—even though the intensity of the latter may be several thousand times greater at the solar surface. Depending upon the velocity vector and its mass, a particle could be forced into a given trajectory. Auroral primaries would be those charges within the requisite energy range, etc., which are allowed to strike the earth's atmosphere and cause aurorae. Obviously it is possible for the solar field to deflect ejecta away from the earth even though on leaving the sun the stream were aimed directly at this planet.

If STÖRMER's trajectories are accepted, an isolated charge leaving the visible photosphere may travel directly to the sunlit or dark side of the earth depending upon the particle's magnetic rigidity. Similarly, for the particles emitted from behind the sun. If the CHAPMAN and FERRARO theory is considered, then it is conceivable that particles having some given energy range may be forced into a solar ring current located at some distance outwards from the sun. This ring current may build up relatively large particle concentrations. Once within this current, the particles may be trapped and allowed to dribble out towards the earth days later, again causing aurorae and magnetic effects not associated with visible solar disturbances. These factors may be of importance in causing the geomagnetic disturbances attributed to *M* regions. If the general solar field varies markedly with time, or if different mass particles were emitted on different occasions, the effects would be similar but somewhat more complicated.

It might be mentioned that it is possible to form a high energy beam within a stream of particles leaving the sun by a mass-spectrometer or focusing action. In this event the narrow beam of high energy corpuscles would exist within the large cross section of the low energy particle stream. Thus, even though the stream might engulf the earth, the high energy particles may completely miss it. Because of their energies the high energy particles, identified as cosmic ray primaries, may travel directly from the visible photosphere to the sunlit earth. However, the slower auroral primaries may be forced into rather complicated paths which eventually lead them to the earth's dark or light side several days later. Several instances of this type have already been noticed.

Geophysics Research Directorate, N. C. GERSON
AF Cambridge Research Center, Cambridge, Mass.

VOL. 4 1953

Journal of Atmospheric and Terrestrial Physics, 1954, Vol. 4, pp. 343 to 345. Pergamon Press Ltd., London

Remarks on the article of J. Clay: "Ions and condensation nuclei in the atmosphere, balance of ions and value of cosmic radiation at sea level"

VICTOR F. HESS

Fordham University, New York, N.Y.

(*Received* 28 *September* 1953)

In the article quoted above (*Journal of Atmospheric and Terrestrial Physics*, Vol. 3, No. 3, April 1953) the author has made statements which cannot be left unchallenged.

1. He asserts that R. V. VANCOUR and I (1950) had said "that no countings were available of condensation nuclei over the ocean." There is no such statement in our paper. Even before our paper was published I made such determinations over the North Atlantic in the summer of 1948 myself (1948) and quoted previous determinations of WIGAND, PARKINSON, MAURAIN and DEVAUX, and LANDSBERG, between 1928 and 1938, omitting determinations of CLAY and his collaborators since these were partly made at locations not very far from the European, African, and Asian continents.

2. J. CLAY says in his paper: "Our measurements were made in order to strike the ionization balance on the ocean. This balance cannot be checked if the sum total of ions/cc present in the atmosphere is to be taken as the basic value, because this sum total can never be ascertained with sufficient reliability. The ions of lowest mobility are for the greater part not of recent production and their number no longer depends on the sources of ionization alone. . . ." I think Dr. CLAY is entirely wrong in bringing this argument into his discussion of the ionization balance of the atmosphere.

The meaning of the word "ionization balance of the atmosphere" is clear enough; we want to know if the number of small ions actually observed can be accounted for, quantitatively, if the rate of ion production is known and a reliable form of the law of recombination of ions is used.

For the free atmosphere SCHWEIDLER's linear law of recombination has been found to hold, at least approximately, even if the number of nuclei is rather low (1000 per cm^3), as it is over the oceans. If the rate of production of small ions is q (ion pairs per cm^3 sec) and θ is the mean life of the small ions, then the product

$$q\theta = n$$

should give n the number of small ions per cm^3 present in the free atmosphere.

This equation has been tested by many authors and the procedure ordinarily followed was a determination of q and of θ (by one of the two methods suggested by E. SCHWEIDLER) and then a comparison of the product with the values of n as obtained by direct measurement with an Ebert ion counter (aspiration method).

The total number of ions (small and large) does not enter directly in this formula. The real difficulty in testing the Schweidler recombination law is that both q and θ can only be measured in closed metallic chambers whereas n as measured with the aspiration method, pertains more or less accurately to the conditions in the free atmosphere.

If in a typical aspiration ion counter air is drawn through the cylindrical condenser (cross section 28 cm^2) at a rate of 5000 cm^3 per sec and the length of the inner electrode is about 20 cm, the ions collected per second from the outer air drawn into the tube were contained in 5000 cm^3 while at the same time (in one second) ions produced in the immediate vicinity of the collecting electrode (500 cm^3) are also measured. Thus, 90% of number of small ions counted originated in the free atmosphere.

The measurement of q cannot be made in a vessel without metallic walls and therefore the result may be vitiated by residual ions formed by radiations from the walls.

Furthermore, the measurement of the mean life of small ions by SCHWEIDLER's method II involves a measurement of the characteristic of the current in an ionization chamber which

naturally takes quite some time (30 min for a complete curve), enough to alter conditions by depositing large ions during the whole period of the measurements. Therefore, we cannot expect perfect confirmation of the equation on account of these discrepancies inherent in the method *V*. The use of SCHWEIDLER's "method I" would perhaps be more advantageous.

J. J. NOLAN and P. J. NOLAN were well aware of these difficulties and proposed several other forms of the recombination law in recent years. G. R. WAIT (1940) suggested a formula $q = an + bnN$ where N is the number of large ions of one sign which is similar to J. J. NOLAN's (1940) expression $q = an + bnZ$ (Z denoting the total number of nuclei).

From all these expressions, it is clear that the total number of ions ($n + N$) enters nowhere and that only the mean lifetime of small ions is influenced by the nuclei or the large ions.

3. Dr. CLAY further says in his paper: "In his calculations for the ion balance on Helgoland, HESS borrowed the value of 500 fast ions per cm^3 in the atmosphere from the Carnegie expedition. However, this value lies far above that for the newly produced ions, as observed by us."

Although I believe that nobody will blame me for "borrowing" the value of 500 ions/cm^3 from the Carnegie expedition, I wish to emphasize that the mean value of the Carnegie measurements over sea (1915–1921) was 550 and that I preferred to reduce this to an approximate figure of 500 since it was not unlikely that part of the large ions were caught also and counted as small ions in these older observations. If the voltage in the aspirating condenser is set so as to collect all small ions (mobility above 1 cm/sec : volt/cm) the number of intermediate and large ions caught will be negligible. This fraction can be determined if a careful analysis of the current *vs.* voltage is made according to H. ISRAEL's (1931) method.

However it is very difficult to reconcile J. CLAY's results (mean value for the number of small ions over sea 213 per cm^3) with the ones of the Carnegie expeditions (500). It must be kept in mind that in general disturbing factors tend to produce rather too small values of n. Therefore I would hesitate to accept CLAY's figures as truly representative.

As I have pointed out in a paper written in 1927 SCHWEIDLER's method II for the determination of the mean life of the small ions can also be used to derive a value for the conductivity of the air and of the number of small ions inside of the chamber. In using this method for n and taking the values for q from his cosmic ray measurements he finds this the "simplest method of investigating the ionization balance." Besides, he says, "the underlying value is here defined better than the total number of fast ions present in the atmosphere can ever be."

Here again the fact is overlooked that neither the conductivity nor the number of small ions as derived from SCHWEIDLER's theory represents truly the respective values in the free atmosphere.

Dr. CLAY (Table 4) points out that the polar conductivity as obtained with the aspiration method is consistently lower than the one derived from the SCHWEIDLER (method II) and "The disparity in the results of the methods grows smaller in the direction of the higher mobility limit which means that there are always ions with much greater mobility than the one the apparatus has been tuned to."

This statement is hard to understand. I have never noticed that the conductivities (aspiration method) depend on the limiting mobility chosen, as long as not all small ions are collected.

Moreover ions of mobilities of 20 cm/sec : Volt/cm do not exist, at atmospheric pressures. Their presence would be clearly revealed by kinks in the characteristic of an aspiration apparatus (Gerdien or Ebert type) if the current (number of ions collected per sec) is plotted versus the voltage across the condenser. Recently such observations were made over the North Atlantic between New York and Liverpool (HESS and PARKINSON, 1952; PARKINSON and WELLER, 1953) and no indication was found for the presence of small ions with mobilities above 2 cm/sec: Volt/cm.

It is not surprising that in general the conductivity as derived from SCHWEIDLER's recombination method, in a closed vessel comes out somewhat higher than the conductivity outside actually is, on account of secondary radiation produced in the vessel.

In evaluating the life time of small ions (using SCHWEIDLER's method II) it is well to remember that in SCHWEIDLER's formula the sum of the mobilities of the small ions enters while F. J. W. WHIPPLE (1933) has shown that only the mobility of the ions of one sign should be used. Thus the lifetime derived with WHIPPLE's correction is twice as great as those derived from the

VOL. 4 1953

original SCHWEIDLER (method II) formula. Unfortunately, it is not stated in CLAY'S paper if he used WHIPPLE'S correction in computing the mean life of the small ions. I assume that the corrected formula has been used.

5. The discrepancy between the values for the ionization by cosmic rays (without secondaries) at sea level obtained by CLAY and his collaborators (1·66 I) and myself (1·88 I) remains unexplained.

A clarification could be obtained if measurements of this sort were made with the same apparatus in Amsterdam and New York.

I wish to apologize to Dr. CLAY for having omitted reference to some papers by himself, by VAN ALPHEN, H. F. JONGEN and P. H. CLAY which were published shortly before World War II and in 1942, and were not accessible for me at that time.

REFERENCES

HESS, V. F.	1927	Sitz. Ber. Akad. d. Wiss, Wien IIa **136,** 603
HESS, V. F.	1948	Terr. Mag. **53,** 399
HESS, V. F., and PARKINSON, W. D.	1952	Fordham University, U.S. Air Force Research Project Af19 (122)–409, Scientific Report No. 2, "Small Ion Balance Over the Ocean" (December), p. 10–11, Fig. 4
HESS, V. F., and VANCOUR, R. V.	1950	J. Atmosph. Terr. Phys. **1,** 13
ISRAEL, H.	1931	Gerlands Beitr. **31,** 173
NOLAN, J. J.	1940	Proc. Roy. Irish Acad. **A46,** 77
PARKINSON, W. D., and WELLER, R. I.	1953	J. Geophys. Res. **58,** 270
WAIT, G. R.	1940	Terr. Magn. **40,** 209
WHIPPLE, F. J. W.	1933	Proc. Phys. Soc. London **45,** 367

DL.
4
953

Journal of Atmospheric and Terrestrial Physics, 1954, Vol. 4, pp. 346 to 348. Pergamon Press Ltd., London

Indices of geomagnetic activity

of the Observatories ABINGER (*Ab*), ESKDALEMUIR (*Es*) and LERWICK (*Le*)

September to November 1953

The figures given on pages 346 to 348 represent the K-indices for three-hour intervals, beginning with 00–03 hrs for the first and ending with 21–24 hrs for the eighth figure.

Day	*Ab* Range for $K = 9 : 500\gamma$			*Es* Range for $K = 9 : 750\gamma$			*Le* Range for $K = 9 : 1000\gamma$			*Day*
	K-Indices		*Sum*	*K-Indices*		*Sum*	*K-Indices*		*Sum*	
1	2333	4432	24	2232	3332	20	2122	3332	18	1
2	5332	3444	28	4332	2343	24	4322	2343	23	2
3	2233	3666	31	2222	3656	28	3222	3548	29	3
4	6544	3454	35	6633	3354	33	7834	3357	40	4
5	5423	3243	26	5422	3243	25	6422	2143	24	5
6	0143	3133	18	1143	3233	20	0142	3133	17	6
7	2233	2413	20	2233	2423	21	3232	2323	20	7
8	3223	2321	18	3223	2221	17	3123	2221	16	8
9	3322	2112	16	3312	2212	16	3311	1211	13	9
10	1122	3324	18	1022	2324	16	2122	2224	17	10
11	5233	3122	21	5222	3222	20	5322	3121	19	11
12	2123	3333	20	2123	3333	20	2122	2233	17	12
13	3233	2223	20	3233	2213	19	3223	2113	17	13
14	0122	2211	11	0110	2110	6	1011	1110	6	14
15	0233	4453	24	1223	4352	22	0223	3352	20	15
16	3233	3443	25	1223	3432	20	2212	2332	17	16
17	3332	4342	24	3231	3242	20	3322	3242	21	17
18	2231	1345	21	2231	1345	21	2121	1246	19	18
19	5455	5655	40	5545	4655	39	7854	5766	48	19
20	5333	4455	32	4333	3444	28	6432	4454	32	20
21	4343	4534	30	4333	4434	28	5433	3433	28	21
22	4443	3345	30	4443	3244	28	6443	2144	28	22
23	4444	5554	35	3444	4444	31	4443	5454	33	23
24	4354	4353	31	4343	4353	29	4343	4342	27	24
25	2133	4342	22	2122	4332	19	2222	3332	19	25
26	1122	2334	18	1222	3234	19	1112	2224	15	26
27	5433	3232	25	4433	2232	23	4432	3232	23	27
28	3322	2232	19	3321	2231	17	3221	2121	14	28
29	1112	3221	13	1211	2211	11	1111	2110	8	29
30	0223	2343	19	0222	2233	16	0111	2232	12	30

VOL. 4 1953

October 1953

Day	*Ab* Range for $K = 9 : 500\gamma$		*Sum*	*Es* Range for $K = 9 : 750\gamma$		*Sum*	*Le* Range for $K = 9 : 1000\gamma$		*Sum*	*Day*
	K-Indices			*K-Indices*			*K-Indices*			
1	3423	3243	24	2422	3242	21	2421	2232	18	1
2	1212	2331	15	1022	2332	15	1111	1221	10	2
3	1213	2212	14	1222	2212	14	0211	1211	9	3
4	2102	1232	13	2102	2232	14	2111	1132	12	4
5	1123	1112	12	1121	2121	11	1111	0011	6	5
6	2122	3122	15	2122	3122	15	2111	2111	10	6
7	2043	3213	18	2133	3213	18	1032	3113	14	7
8	1233	3334	22	1233	3224	20	1222	2214	16	8
9	2333	2112	17	2232	2222	17	2221	2112	13	9
10	3212	2343	20	2111	2243	16	3211	2243	18	10
11	1112	3331	15	1112	3332	16	1111	2331	13	11
12	1012	1132	11	1001	1132	9	0011	1021	6	12
13	1113	2222	14	1102	1222	11	0111	1222	10	13
14	1122	1211	11	1111	1100	6	1110	1110	6	14
15	0122	5655	26	0022	4765	26	0011	4776	26	15
16	5333	3656	34	5332	3556	32	6332	2467	33	16
17	4334	4654	33	4324	3553	29	6223	3554	30	13
18	3343	4554	31	3443	4554	32	5543	5566	39	18
19	5345	5554	36	5444	4553	34	6544	4543	35	19
20	3443	4555	33	4433	4454	31	4433	4544	31	20
21	3245	4312	24	3244	4211	21	3234	3311	20	21
22	1433	3543	26	1333	3432	22	1333	2433	22	22
23	3333	3331	22	3322	2332	20	2322	2331	18	23
24	1122	4212	15	0123	4222	16	1012	3202	11	24
25	3323	2123	19	2322	3212	17	2222	2112	14	25
26	2123	1133	16	2122	2122	14	2112	1022	11	26
27	4333	3354	28	4332	2354	26	4232	2353	24	27
28	3222	1233	18	3221	1233	17	3311	0233	16	28
29	3312	3334	22	2322	3334	22	2211	2244	18	29
30	3223	2231	18	3222	2121	15	3111	2121	12	30
31	3222	2100	12	3221	2110	12	3221	2101	12	31

November 1953

1	0012	1232	11	0012	1132	10	0111	1132	10	1
2	1122	1211	11	1012	0112	8	1011	0111	6	2
3	1111	2222	12	1011	2322	12	1112	1322	13	3
4	3122	1112	13	2211	1112	11	2111	1002	8	4
5	2332	3331	20	2332	3330	19	2322	3330	18	5
6	2122	3111	13	2122	3110	12	2112	2101	10	6
7	0122	2131	12	0121	1121	9	0111	1122	9	7
8	0011	1113	8	0021	1213	10	1011	1213	10	8
9	3111	1112	11	3111	0002	8	3110	0001	6	9
10	0012	1120	7	0010	0000	1	0000	0011	2	10
11	0011	2132	10	0001	2232	10	0000	1033	7	11
12	3122	3355	24	3212	3245	22	3201	3245	20	12
13	4434	5454	33	4433	5444	31	4333	4445	30	13
14	4523	5544	32	3422	5544	30	3422	4554	29	14

K-Indices

November 1953 (*contd.*)

Day	*Ab* *Range for K = 9 : 500γ*			*Es* *Range for K = 9 : 750γ*			*Le* *Range for K = 9 : 1000γ*			*Day*
	K-Indices		*Sum*	*K-Indices*		*Sum*	*K-Indices*		*Sum*	
15	2354	4553	31	2344	4553	30	1334	4552	27	15
16	4433	3554	31	4432	3443	27	5532	3444	30	16
17	3222	3445	25	3222	3444	24	3312	2434	22	17
18	3324	3333	24	3323	2323	21	3323	2313	20	18
19	4433	5355	32	4433	4354	30	3433	4354	29	19
20	3443	3454	30	3333	3344	26	3333	2244	24	20
21	3223	3132	19	3223	3132	19	3122	2133	17	21
22	1011	2132	11	1021	2022	10	1011	1022	8	22
23	2133	3465	27	2133	3454	25	1123	3365	24	23
24	3333	1223	20	3232	1223	18	4232	1222	18	24
25	2332	2234	21	2322	2224	19	1321	2124	16	25
26	3222	1121	14	3111	1121	11	3111	0012	9	26
27	2132	2222	16	2122	2222	15	2121	2212	13	27
28	2111	0012	8	1101	0002	5	1110	0002	5	28
29	2212	1001	9	2112	1001	8	2211	0001	7	29
30	1011	2121	9	1010	1121	7	1011	1111	7	30

VOL.
4
1953

CORRECTION

In "The reception of atmospherics at high frequencies" by F. HORNER, published in Vol. 4, No. 3 of this journal (p. 129) a regrettable error occurred during production. The diagrams above the captions to Figures 8 and 9 were inadvertently transposed. The diagram which appears above the caption to Figure 8 should appear above the caption to Figure 9 and *vice versa*.

OL.
4
953